Cambridge Healthtech Institute's 2nd Annual

Bio-IT World EUROPE CONFERENCE & EXPO 2010

5–7 October 2010 • Exhibition Grounds • Hannover, Germany

5-6 October

IT Infrastructure & Informatics:
Data Storage, Analysis and Visualization

NGS Data Management:
Sequencing Systems, Storage, and Analysis

6-7 October

IT Infrastructure & Informatics:
Collaborative Software, Standards & Web Technologies

Data Integration & Knowledge Management

Premier Sponsors:

BIOTEAM Enabling Science

hp

life technologies™

applied biosystems | invitrogen

Corporate Support Sponsor:

panasas

Register by **16 July** and **Save** up to **€250!**

Held in Conjunction with

BIO TECHNICA

Europe's No.1 Event in Biotechnology and Life Sciences

Organized by:

Cambridge Healthtech Institute

250 First Avenue
Suite 300
Needham, MA 02494
Phone: +1.781.972.5400
Toll-free in the US:
888.999.6288
Fax: +1 781.972.5425

PRE-CONFERENCE SHORT COURSES*

Monday, 4 October

10:00 – 13:30

Making Sense of Next-Gen Sequencing Data

Organized by BIOTEAM Enabling Science

With the widespread deployment of second-generation sequencing and the emergence of new third-generation sequencing platforms, the extraordinary throughput of next-generation sequencing (NGS) technology is outpacing our ability to analyze and interpret the data. Researchers need productive strategies to cope and handle this deluge of data. This workshop will focus on practical informatics methods, strategies and software tools for transforming NGS data into usable information.

What you will learn:

- NGS assembly and annotation methods
- Tools for data analysis
- An appraisal of commercial software and freeware
- Case studies of analysis to support research

Short Course Instructors:

Instructors to be announced.

14:00 – 17:00

Creating Synergy – Introduction to Biomedical Data Fusion

Systems biology and personalized medicine increasingly require a synergistic consideration of different molecular or clinical data sets. Making such heterogeneous data available is only the first step for obtaining the big picture through a coherent analysis, i.e. data fusion. This introductory tutorial will provide a broad overview of the different options and methodologies for making the most of your data through data fusion.

- A principled approach to data fusion
- Powerful methods from machine learning, multivariate statistics and pattern recognition
- How to deal with any kind of data
- QTL mapping of omics data
- Application examples in cancer and diabetes

Short Course Instructors:

Juergen von Frese, Ph.D., Managing Director, Data Analysis Solutions DASOL GmbH
Marc-Emmanuel Dumas, Ph.D., Lecturer in Systems Biomedicine, Imperial College
Gunnar Rätsch, Ph.D., Friedrich Miescher Laboratory, Max Planck Society

Who Should Attend:

Researchers with a basic understanding of omics data analysis who want to combine data from different sources for extracting maximal information.

14:00 – 17:00

Cloud Computing for Life Sciences

Organized by CYCLE COMPUTING

Cycle Computing is leading the efforts for many life science organizations in using the cloud, helping research labs and companies leverage internal and external clouds for collaboration, calculations, and storage. We'll cover real world use cases across drug discovery & design, collaboration, next generation sequencing, proteomics, software as a service, and bioinformatics, to explore how life sciences are using cloud computing, its challenges and effectiveness, how money can be saved by an organization, and regulatory compliance. Join thought leaders in this day long workshop to examine how cloud computing can be used effectively as an external IT service and an internal computing model.

Short Course Instructors:

Instructors to be announced.

14:00 – 17:00

Visualization of Large-Scale Biological Data

Data visualization has become increasingly important for life scientists as the amount of data generated in biomedical studies continues to grow rapidly. Visual representations are powerful tools in exploring large quantities of data quickly, helping to detect patterns and generate hypotheses, which guide further analyses. This practical course will provide a comprehensive view on utilizing visualization to support the analysis of large biological data sets, and will cover interaction networks and biochemical pathways, as well as transcriptomics, proteomics and metabolomics data.

- Visualization principles and pitfalls
- The roles of visualization in data analysis
- Key methods and software tools
- Integration of visualization with automated methods: Visual Analytics
- Future technologies

Short Course Instructor:

Nils Gehlenborg, European Bioinformatics Institute, Wellcome Trust Genome Campus
Kay Nieselt, Ph.D., Group Leader, Center for Bioinformatics Tuebingen, University of Tuebingen

**Separate Registration Required.*

MEDIA PARTNERS

Official Publication: Bio-IT World

Lead Sponsoring Publications: eCliniqua, GEN Genetic Engineering & Biotechnology News, nature, Science AAAS, The Scientist

Sponsoring Publications: ddn Drug Discovery & Development, Genome Technology, PharmaVOICE

Web Partners: BioSpace.com, genomeweb, hum-molgen.org central gateway to HUMan MOLecular GENetics, www.PharmCast.com – Pharmaceutical Internet –, SelectScience.net

IT Infrastructure & Informatics:
Data Storage, Analysis and Visualization
5-6 October 2010

TUESDAY, 5 OCTOBER

9:00 Conference Registration and Morning Coffee

LARGE SCALE DATA STORAGE, RETRIEVAL AND ACCESSIBILITY – LEADING EU CENTERS

9:30 Chairperson's Opening Remarks *Sponsored by* panasas

Derek Burke, Director, International Marketing, Panasas

9:35 Next-Generation Sequencing Data: Next-Generation Management Problems

Guy Coates, Ph.D., Informatics System Group, The Wellcome Trust Sanger

Sequencing centers are now collecting and storing large amounts of data from their sequencing machines. However, as datasets get ever larger, simply keeping track of the data becomes a challenge; datasets are lost or needlessly duplicated. The situation is further complicated as datasets become dispersed over collaborative and cloud infrastructures. The talk will explore some of the technologies we are using to help people manage and share their data in the next-gen sequencing era.

10:05 Keeping Pace—Scalable IT Infrastructures to Support Data Intensive Science

Rupert Lueck, Ph.D., Head, IT Services, EMBL Heidelberg

The IT infrastructure required to support science at EMBL is seriously challenged by the enormous amounts of data generated by technologies such as NGS and high-throughput microscopy that are used in large-scale and interdisciplinary systems biology projects. This talk highlights our strategies to ensure a scalable, reliable and cost efficient IT environment to keep pace with the rapidly growing demands for high performance storage and compute power.

10:35 Coffee Break

11:00 Omics Data - Serving Sequencing at the National Level

Francois Artiguenave, Ph.D., Head, Bioinformatics Laboratory, Genoscope

Life science has been profoundly impacted by technological advances allowing faster and cheaper DNA sequencing. Opening a wide range of applications, the last sequencing platforms raised new challenges in processing, analysing and interpreting massive data. The growing role of informatics and bioinformatics will be illustrated by providing some figures about genome sequencing and others applications aimed at unraveling biological mechanisms.

11:30 Speaker to be Announced

12:00 Presentation *Sponsored by* hp

12:30 Lunch for Purchase in the Exhibit Hall

13:45 Dedicated Poster Viewing in the Exhibit Hall

STORAGE: SOLUTIONS FOR MASSIVE SEQUENCING DATA MANAGEMENT

14:30 Chairperson's Remarks

14:35 Featured Presentation: Sequencing Data Storage *Sponsored by* BIOTEAM Enabling Science

Chris Dagdigian, Founding Partner & Director, Technology, BioTeam, Inc.

Next-Generation Sequencing (NGS) instruments are forcing evolutionary and revolutionary changes in research IT architectures & infrastructures. Chemistry and lab protocols are advancing faster than the underlying IT systems and methods, leading to a crisis of capability in many organizations. This presentation will focus on "science-centric storage" for life science informatics, with specific attention on requirements, trends, data management methods and emerging best practices.

15:05 Experiences from the European Bioinformatics Institute's Data Resources, Storage, and Management

Guy Cochrane, Ph.D., European Nucleotide Archive Team Leader, Bioinformatics Institute

15:35 Refreshment Break

16:00 Sponsored Presentation *(Opportunity Available)*

16:30 How to Overcome the 100 Miles between Petabases and Petabytes

Jürgen Eils, Bioinformatics Database Group Leader, German Cancer Research Center

Recently, Heidelberg University received a grant to build the largest data storage facility in Germany at 5-10 petabytes. From a management and logicistical perspective, the massive throughput of next-generation sequencing requires new concepts and strategies. One problem is the long distance transport of data from the sequencer machine to the data storage facility. We will present strategies and concepts with emphasis on reusability and sustainability for storing and retrieving the comprehensive collection of sequencing data in combination with associated clinical and histopathological annotation data – all in accordance with the International Cancer Genome Consortium (ICGC) guidelines.

17:00 Enhanced Scalabilty, Large Data Volumes Management, Integrated Analysis, and NGS Informatics Support in a Medical Setting

Andrew Stubbs, Ph.D., Assistant Professor, Department of Bioinformatics, Erasmus Medical Center

17:30 The First Success Stories after the Swedish Buildup of Computational Power and Large-Scale Storage for Gene-Sequence Data

Ingela Nystrom, Ph.D., Director, UPPMAX, Center for Image Analysis, Uppsala University

Last year, we reported on the buildup of a system at Uppsala University, Sweden, intended for researchers who deal with the large-scale data from modern gene-sequencing technology. The system has 1200 cores, 4 TB RAM, and 500 TB storage. Now, we report our first success stories, e.g., the whole-genome resequencing project revealing signatures of selection during chicken domestication.

18:15 Interactive Breakout Discussion Groups:

Storage of Omic Data - The Cloud & Beyond

Moderator: Chris Dagdigian, Founding Partner & Director of Technology, BioTeam, Inc.

- Science-centric storage
- Data management issues
- Best practices

Semantic Web

Moderator: To be Announced

- Increasing accessibility of data to the user
- Linked data framework
- Dealing with scale using semantic web

Drug Discovery Informatics –Knowledge Integration

Moderator: To be Announced

- Sharing data inside and out
- Open source
- Knowledge management accelerating discovery
- Translational medicine

19:15 – 21:00 CHI Networking Reception

WEDNESDAY, 6 OCTOBER

9:00 Conference Registration and Morning Coffee

IMAGING SOLUTIONS IN DRUG R&D: DATA ANALYSIS, IMPLEMENTATION, STANDARDS AND QUALITY

9:30 Chairperson's Opening Remarks

9:35 Implementation & Use of PHAEDRA, a Standards-Based System for High-Content Image Analysis and Evaluation

Frans Cornelissen, IT Manager, Janssen Pharmaceutical

High-content Image Analysis based screening (HTS-HCA) technology is an important drug discovery tool for identification of biological probes and drug leads by screening large volume, diverse biochemical and cell-based assays, using image capture & analysis. The architecture and usage of the PHAEDRA environment will be described, using examples like measurement of 3D tumor colony size on brightfield image stacks, quantification of dendritic length, spine density and spine diameter in 3D fluorescent image stacks.

9:55 Lessons Learned in Imaging Informatics for Drug Discovery

Gudrun Zahlmann, Ph.D., Manager, Imaging Infrastructure, pRED, F. Hoffman-La Roche Ltd.

10:15 Quantitative Image Analysis Tools for Biological Research

Ewert Bengtsson, Ph.D., Professor, Center for Image Analysis, Uppsala University

Pure visual analysis of microscopy images is limited in its ability to provide objective quantitative information. Here computerized image analysis can provide automated, quantitative tools enabling accurate and high throughput analysis. We will describe methods for improved microscope image analysis ranging from better utilization of the color information via robust modeling and segmentation to quantification and classification methods.

10:35 Coffee Break

11:00 Sponsored Presentation *(Opportunity Available)*

KEYNOTE PRESENTATIONS - IT & INFORMATICS IN SUPPORT OF DRUG DISCOVERY

11:30 Next-Generation Interfaces and Interaction with Complex Information Landscapes

Bryn Roberts, Ph.D., Global Head, Informatics, Pharma Research and Early Development, F. Hoffmann-La Roche AG

Presentation of background and proof-of-concept projects on two main themes: Potential approaches of enabling scientists to navigate complex, heterogeneous information using semantic integration technologies and a next generation of user interfaces; and Enabling teams to generate, explore and progress hypotheses together using collaborative computer interfaces.

12:00 IT Support of Drug Discovery

Jacob deVlieg, Ph.D., Professor & Global Head, Molecular Design and Informatics, Merck

12:30 Lunch for Purchase in the Exhibit Hall

13:00 Dedicated Poster Viewing in the Exhibit Hall

13:30 Close of Conference

TRACK 2

NGS Data Management:
Sequencing Systems, Storage, and Analysis
5-6 October 2010

TUESDAY, 5 OCTOBER

9:00 Conference Registration and Morning Coffee

INFORMATICS: MEETING THE CHALLENGE OF TURNING NGS DATA INTO KNOWLEDGE

9:30 Chairperson's Opening Remarks

9:35 Into the Unknown: Expression Profiling without Genome Sequence Information by Next-Generation Sequencing

Fabian Birzele, Ph.D., Genomics, Boehringer Ingelheim Pharma, GmbH & Co. KG

Expression profiling in organisms lacking genome or transcriptome sequence information is feasible by combining Illumina's mRNA-seq technology with a novel bioinformatics pipeline that integrates assembled and annotated sequences from read data with information derived from related organisms. Using the Chinese hamster as a model, expression patterns for more than 13000 genes can be analyzed. A detailed analysis of selected biological functions such as DNA replication and cell cycle control demonstrate the potential of NGS expression profiling in organisms without extended genome sequence to improve data quantity and quality.

10:05 Iterative Read Mapping and Assembly Allows the Use of a More Distant Reference in Metagenome Assembly

Bas E. Dutilh, Ph.D., Centre for Molecular and Biomolecular Informatics, Nijmegen Center for Molecular Life Sciences, Radboud University Nijmegen Medical Center

Using a reference can greatly improve the assembly of next-generation sequencing reads, but a closely related genome is not always available; however, by using a permissive mapping algorithm and by iterating the mapping and assembly several times, a more distant reference can still be used. From a short-read sequenced enriched bioreactor, we construct a sequence that captures the consensus of the population's metagenome.

10:35 Coffee Break

11:00 Utilizing Next-Generation Sequencing to Analyze the Complex Genome of Barley

Burkhard Steuernagel, Department of Cytogenetics and Genome Analysis, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)

De novo sequencing the entire genome of a large complex plant genome like the one of barley (*Hordeum vulgare L.*) is a major challenge both in terms of experimental feasibility and costs. The emergence and breathtaking progress of next-generation sequencing technologies has put this goal into focus. As a proof of concept, 3500 clones were sequenced to at least 15-fold coverage in pools of up to 48 barcoded BACs. The high quality of the assembled contigs show the application in a clone based sequencing strategy for sequencing the barley genome.

11:30 Fully Automated Genome Annotation with Deep RNA Sequencing

Gunnar Rätsch, Ph.D., Group Leader, Friedrich Miescher Laboratory, Max Planck Society

The development of high-throughput sequencing technologies allows the determination of the complete set of RNA-transcripts expressed under a given condition. We present accurate and efficient computational methods to automatically annotate genes and transcripts together with their expression levels from deep RNA sequencing. We illustrate that these approaches only require RNA deep sequencing reads and the genome sequence and are able to determine very accurate genome annotations. Hence, high quality genome annotation can now be fully automated. We applied these techniques to annotate novel and to reannotate well-characterized genomes and could identify many yet undiscovered transcripts.

12:00 Sponsored Presentation *(Opportunities Available)*

12:30 Lunch for Purchase in the Exhibit Hall

13:45 Dedicated Poster Viewing in the Exhibit Hall

STORAGE: SOLUTIONS FOR MASSIVE SEQUENCING DATA MANAGEMENT

14:30 Chairperson's Remarks

14:35 Featured Presentation: Sequencing Data Storage *Sponsored by* BIOTEAM Enabling Science

Chris Dagdigian, Founding Partner & Director of Technology, BioTeam, Inc.

Next-Generation Sequencing (NGS) instruments are forcing evolutionary and revolutionary changes in research IT architectures & infrastructures. Chemistry and lab protocols are advancing faster than the underlying IT systems and methods leading to a crisis of capability in many organizations. This presentation will focus on "science-centric storage" for life science informatics with specific attention on requirements, trends, data management methods and emerging best practices.

15:05 Experiences from the European Bioinformatics Institute's Data Resources, Storage, and Management

Guy Cochrane, Ph.D., European Nucleotide Archive Team Leader, Bioinformatics Institute

15:35 Refreshment Break

16:00 Sponsored Presentation *(Opportunity Available)*

16:30 How to Overcome the 100 Miles between Petabases and Petabytes

Jürgen Eils, Bioinformatics Database Group Leader, German Cancer Research Center

Recently, Heidelberg University received a grant to build the largest data storage facility in Germany at 5-10 petabytes. From a management and logical perspective, the massive throughput of next generation sequencing requires new concepts and strategies. One problem is the long distance transport of data from the sequencer machine to the data storage facility. We will present strategies and concepts with emphasis on reusability and sustainability for storing and retrieving the comprehensive collection of sequencing data in combination with associated clinical and histopathological annotation data – all in accordance with the International Cancer Genome Consortium (ICGC) guidelines.

17:00 Enhanced Scalability, Large Data Volumes Management, Integrated Analysis, and NGS Informatics Support in a Medical Setting

Andrew Stubbs, Ph.D., Assistant Professor, Department of Bioinformatics, Erasmus Medical Center

17:30 The First Success Stories after the Swedish Buildup of Computational Power and Large-Scale Storage for Gene-Sequence Data

Ingela Nystrom, Ph.D., Director, UPPMAX, Center for Imaging Analysis, Uppsala University

Last year, we reported on the buildup of a system at Uppsala University, Sweden, intended for researchers who deal with the large-scale data from modern gene-sequencing technology. The system has 1200 cores, 4 TB RAM, and 500 TB storage. Now, we report our first success stories, e.g., the whole-genome resequencing project revealing signatures of selection during chicken domestication.

18:15 Interactive Breakout Discussion Groups

Please see Page 3 for Details.

19:15 – 21:00 CHI Networking Reception

WEDNESDAY, 6 OCTOBER

9:00 Conference Registration and Morning Coffee

SEQUENCING DATA: PROVIDES A PROMISING TOOL FOR MOLECULAR DIAGNOSTICS

9:30 Chairperson's Opening Remarks

Kevin Davis, Ph.D., Editor-in-Chief, Bio-IT World

PLENARY KEYNOTES

9:35 Applications of Next-Generation Sequencing in an Academic Medical Hospital: From Single Molecules to a Complete Female Genome

Johan den Dunnen, Ph.D., Head, Leiden Genome Technology Ctr (LGTC), Human and Clinical Genetics, Leiden University Medical Ctr

Medical application of full human genome sequencing to resolve a health problem will soon be a realistic option. Within our academic hospital we have started to implement next-generation sequencing with applications covering many subjects: candidate disease genes, bacterial genomes, single molecules sequencing and the sequencing of a complete female genome. We experienced that technologically, human genome sequencing is feasible, but computationally it was at the limits of our possibilities. The problem resides in interpretation, where efficient analytical tools are largely lacking.

10:05 Drivers, Challenges and Opportunities to Bring Diagnostics Closer to the Points of Need

Rudi Pauwels, Ph.D., Founding Director & CEO, Biocartis

The rapidly expanding atlas of molecular-based biomarkers and the advent of novel technologies are creating new opportunities to improve the outcome for the individual patient by providing tools to implement a more personalized and increasingly more molecular-based medicine. These trends will likely accelerate in a climate of intensive changes and pressures that relate to the various industries, healthcare players and regulatory bodies involved. Although technical, regulatory, economic, business and adoption hurdles need to be overcome, the overall need of affordable healthcare for all is likely to be an important selection pressure, if not critical determinant , for the future evolution of healthcare.

10:35 Coffee Break

11:00 Sponsored Presentation *(Opportunity Available)*

11:30 Bringing Next-Generation Sequencing to the Clinic

Mark P. Stevenson, Ph.D., President and COO, Life Technologies

While the pace of technological improvements has led to lowering of costs, the next frontier will be in demonstrating clinical utility. In select cases, diagnostic tests limited to a panel of genes fail to correctly diagnose all patients who have the clinical manifestations of the disease. We will highlight collaborative work on Charcot-Marie-Tooth and Noonan Syndrome using NGS. In acquired diseases, we are focusing on cancer, where the mutation profiles are so complex that whole genomic approaches are preferred. Key issues around privacy, reimbursement, education, and regulation are essential to realizing the clinical power of NGS.

12:00 PANEL DISCUSSION

Moderator: Kevin Davis, Ph.D., Editor-in-Chief, Bio-IT World

The explosion of next-generation sequencing and other tools for high-throughput genomic analysis is already proving its value, for example in identifying mutations underlying rare Mendelian disorders and rationalizing cancer treatments. But the extraordinary volume and complexity of these data make the informatics of data analysis more costly and time-consuming than generating the original data. As one scientist said, "What use is the $1,000 genome if it costs $20,000 to do the analysis?" In this panel discussion, Drs den Dunnen, Pauwels and Stevenson discuss the challenges of clinically interpreting next-gen genomic data and delivering those data in a timely and accessible fashion to the bedside.

12:30 Lunch for Purchase and Poster Viewing in the Exhibit Hall

13:30 Close of Conference

TRACK 3

IT Infrastructure & Informatics:
Collaborative Software, Standards & Web Technologies
6-7 October 2010

WEDNESDAY, 6 OCTOBER

13:00 Conference Registration

DATA INTEGRATION AND COLLABORATION

14:00 Chairperson's Remarks

14:05 Knowledge Engineering: Integrating the External with the Internal

Ian Dix, Ph.D., Capability Lead for Knowledge Engineering & Information Science, Discovery Information, AstraZeneca

Decision-making within R&D is often compromised by the sheer volume and heterogeneity of decision-relevant biomedical & commercial information available to project teams: project teams struggle to access critical information in the necessary time, at an acceptable quality, coverage, and cost. This challenge is compounded by the formats and accessibility of biomedical information, with the most valuable information being 'locked' within textual formats such as the literature, patents and internal reports. This BioIT presentation will describe AZ's response to this challenge, focusing on external content delivery.

14:35 The Reality of Web Services in the Life Sciences

Carole Goble, Professor of Computer Science, University of Manchester

Finding and using web services to third party online bioinformatics resources is much harder than it should be. This talk will describe the current reality of Web Services in the Life Sciences, and how better practices can be encouraged. BioCatalogue provides a common interface for registering, finding, and monitoring bio-web services and will be described here.

15:05 The New Information Landscape: Towards a New Logic and Practice to Enhance Interoperability and Collaboration in Life Science

Mats Sundgren, Ph.D., Principal Scientist, Global Clinical Development, AstraZeneca R&D

Today the pharmaceutical industry needs to transform itself on many levels to enhance scientific research and innovation output. In order for the pharmaceutical industry to succeed in new research areas like personalized medicine and predictive medicine, one key factor is how to connect their internal information with the external world, including health care and patients, to share and make use of different information types or systems. From a new drug development perspective, the competitive edge is the ability to connect all information domains.

15:35 Refreshment Break

CASE STUDIES OF OPEN COLLABORATIONS AND OPEN SOURCE SOFTWARE

16:00 Sponsored Presentation *(Opportunities Available)*

16:30 Development of a Registration System in a Collaborative Special Interest Group

Jeremy Packer, Ph.D., Head, Bioinformatics, Abbott

17:00 Building Tools for Collaborative and Open Projects in Biology

Lars Jorgensen, Senior Scientific Manager, Production Software, Sequencing Informatics, The Wellcome Trust Sanger Institute

Building tools to support research consortiums and flat research organizations poses a number of challenges. Requirements are often very open ended and the tools are used by a large number of different types of users, from Ph.D. students to dedicated informatics staff. This talk will describe the tools and processes we use to tackle these challenges.

18:30 – 21:00 BIOTECHNICA Night: Beer Hall, Full Dinner Reception, Live Band

(Please register to reserve your complimentary ticket ahead of time. No tickets will be available on-site.)

THURSDAY, 7 OCTOBER

9:00 Conference Registration and Morning Coffee

IMPLEMENTING STANDARDS TO REACH PROCESS EFFICIENCY AND INTEROPERABILITY – METADATA AND BEYOND

9:30 Chairperson's Opening Remarks

9:35 MIBBI: Minimum Reporting Guidelines for Bioscientists

Chris Taylor, Ph.D., Senior Technical Officer, The European Bioinformatics Institute

The last ten years have seen significant progress in various bioscience fields towards standardized guidance for reporting public research, aiming to raise the annotation quality and utility of publicly-available bioscience data. MIBBI integrates the outputs of that (necessarily piecemeal) development, presenting users with a unified set of guideline 'modules' that can easily be assembled and accessed in various forms.

10:05 Can Metadata Repositories Drive Us Closer to Interoperability?

Julia Zhang, Ph.D., Associate Director, Standard & Architecture Program, Genzyme Corporation

Metadata and metadata repositories can bring us closer to interoperability; however, metadata repositories have not matured in their development. Metadata repository requirements and structure will be suggested from this presentation. Technology, system and data integration is a challenge for improving quality and efficiency. Interoperability will be able to help; however, how to reach the interoperability is another challenge. The proposed presentation will address these issues.

10:35 Coffee Break

11:00 Sponsored Presentation *(Opportunity Available)*

11:30 iSee - Dissemination of Structural Biology Data to the Masses

Brian Marsden, Ph.D., P.I., Research Informatics, Structural Genomics Consortium, University of Oxford

Making structural biology data accessible to non-structural biologists is key to ensure that structural information is considered during a chemical probe development or drug discovery project. In collaboration with MolSoft LLC, the SGC has developed an intuitive and interactive platform called iSee that integrates structural biology methods into a 3D document. This platform and our data is freely available online.

ONTOLOGIES AND STATE-OF-THE-ART WEB TECHNOLOGIES

12:00 Semantic Data Integration for Translational Research in Neuroscience

Rudi Verbeeck, Ph.D., Program Manager, Global Information Systems, Johnson & Johnson

To support translational research, we are evaluating the potential of semantic web technology to integrate data from discovery research through to the clinical environment in the area of neuroscience. We present our architecture for data source description and discovery and for query generation. A key component of the architecture is a central ontology that defines common entities used in data source mapping.

12:30 Lunch for Purchase in the Exhibit Hall

13:45 Dedicated Poster Viewing in the Exhibit Hall

ONTOLOGIES AND STATE-OF-THE-ART WEB TECHNOLOGIES CONT'D

14:30 Chairperson's Remarks

14:35 Keeping Track of Data and What the Data Are About: An Exploration in Realism-Based Ontology for Translational Research

Werner Ceusters, Ph.D., Director, Ontology Research Group, NYS Center of Excellence in Bioinformatics & Life Sciences

The Innovative Medicines Initiative (IMI) is a unique pan-European public and private sector collaboration between large and small biopharmaceutical and healthcare companies, regulators, academia and patients. The aim of IMI is to support the faster discovery and development of better medicines for patients and enhance Europe's competitiveness by ensuring that its biopharmaceutical sector remains a dynamic high-technology sector. One specific aim is to create a Knowledge Management Platform to arrive at effective data integration and analysis tools. We will discuss the exact place of ontology in such an endeavor and its relationship to terminologies and information models.

15:05 Sponsored Presentation *(Opportunity Available)*

15:35 Refreshment Break

16:00 Sponsored Presentation *(Opportunity Available)*

16:30 How Can a "Normal End User" Become a Web Service Consumer?

Christian Hauck, Ph.D., Knowledge Management and Competitive Intelligence, Novartis Pharma AG

Web services exist in various variants, but are mostly used by experts to build machine-to-machine interfaces. Without enough consumers, providers don't add value. Without enough providers, consumers get no value. We try to build web service consumer "frontends" for educated scientists and business people to help overcome this "chicken-and-egg" bootstrap problem. Attendees will learn about how to re-use web services that might already exist in their organizations, thereby creating more value out of what is already available.

17:00 Efficient Web-Browsing for the Life Scientist

Reinhard Schneider, Ph.D., Team Leader, Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL)

Anyone who regularly reads life science literature often comes across names of genes, proteins, or small molecules that they would like to know more about. We have developed a new service called Reflect (http://reflect.ws) that can be installed as a plug-in into browsers. Reflect tags gene, protein, and small molecule names in any web page, typically within a few seconds. Clicking on a tagged gene or protein name opens a popup showing a concise summary that includes synonyms, database identifiers, sequence, domains, 3D structure, interaction partners, subcellular location, and related literature. The popups also allow navigation to commonly used databases.

17:30 A Data Warehousing Approach for the Management of Clinical and Operational Information: Specific Challenges within Pharma Development

Norbert Fritz, Ph.D., Development Lead, Business Intelligence Warehouse, Information Management - Product Development Operations, F. Hoffmann-La Roche Ltd

The development of new medicines implies more and more extensive generation and usage of medical information. Management of this broad scope of information requires integration of data generated in heterogeneous data sources and presentation of clinical and operational data in one consolidated data model. Specific challenges for Pharma Development include: evolving and changing medical concepts, disparate systems generating source data, disparate data models and standards within various data sources, transient data quality issues, the need for frequent data refreshes. We will analyze the specific challenges, present approaches to design adequate solutions and discuss real-life experiences.

18:00 Close of Conference

Data Integration & Knowledge Management

6-7 October 2010

WEDNESDAY, 6 OCTOBER

13:00 Conference Registration

DATA INTEGRATION AND COLLABORATION: TEARING DOWN THE WALLS BETWEEN THE KNOWLEDGE SILOS IN THE PHARMACEUTICAL INDUSTRY

14:00 Chairperson's Remarks

14:05 Knowledge Engineering: Integrating the External with the Internal

Ian Dix, Ph.D., Capability Lead for Knowledge Engineering & Information Science, Discovery Information, AstraZeneca

Decision-making within R&D is often compromised by the sheer volume and heterogeneity of decision-relevant biomedical & commercial information available to project teams: project teams struggle to access critical information in the necessary time, at an acceptable quality, coverage, and cost. This challenge is compounded by the formats and accessibility of biomedical information, with the most valuable information being 'locked' within textual formats such as the literature, patents and internal reports. This BioIT presentation will describe AZ's response to this challenge, focusing on external content delivery.

14:35 The Reality of Web Services in the Life Sciences

Carole Goble, Professor of Computer Science, University of Manchester

Finding and using web services to third party online bioinformatics resources is much harder than it should be. This talk will describe the current reality of Web Services in the Life Sciences, and how better practices can be encouraged. BioCatalogue provides a common interface for registering, finding, and monitoring bio-web services and will be described here.

15:05 The New Information Landscape: Towards a New Logic and Practice to Enhance Interoperability and Collaboration in Life Science

Mats Sundgren, Ph.D., Principal Scientist, Global Clinical Development, AstraZeneca R&D

Today the pharmaceutical industry needs to transform itself on many levels to enhance scientific research and innovation output. In order for the pharmaceutical industry to succeed in new research areas like personalized medicine and predictive medicine, one key factor is how to connect their internal information with the external world, including health care and patients, to share and make use of different information types or systems. From a new drug development perspective, the competitive edge is the ability to connect all information domains.

15:35 Refreshment Break

KNOWLEDGE MANAGEMENT FOR TARGET DISCOVERY: ENHANCING BIOLOGICAL KNOWLEDGE BY UTILIZING LARGE SCALE DATASETS

16:00 Sponsored Presentation *(Opportunities Available)*

16:30 Novel Knowledge Integration Approaches to Mining Genome-Wide Data Sets

Mario Albrecht, Ph.D., Research Group Leader, Computational Biology and Applied Algorithmics, Max Planck Institute for Informatics

Novel computational approaches are required to integrate and mine rapidly increasing amounts of molecular data for translational systems research. This talk will present new bioinformatics methods how to transform gene and protein sets into biological knowledge and potential drug targets. Possible solutions to current prioritization challenges of large-scale data sets generated by genome-wide association studies, RNA interference screens, and various detection techniques of protein interactions and complexes will utilize comprehensive network and pathway information.

17:00 *In silico* Target Discovery Based on Flexible Pipelining and Meta-Analysis

James Cai, M.D., Head, Biomedical Informatics, Pharma Research & Early Development (pRED) Informatics, Hoffmann-La Roche

Target discovery relies heavily on the intelligent use of scientific information. Recently high dimensional genomics data have become increasingly critical in discovering new targets and biomarkers. However it is difficult to combine findings from the large number of genomics studies and integrate across multiple data types. This presentation will describe a new *in silico* target discovery approach based on flexible pipelining and quantitative data integration using a meta-analysis framework. Examples will be given to show how results from multiple gene expression studies can be combined with mutation, disease association and competitor information to prioritize potential targets in oncology.

18:30 – 21:00 BIOTECHNICA Night: Beer Hall, Full Dinner Reception, Live Band

THURSDAY, 7 OCTOBER

9:00 Conference Registration and Morning Coffee

TRANSLATIONAL INFORMATICS: KNOWLEDGE FERTILIZATION BETWEEN BENCH AND BEDSIDE

9:30 Chairperson's Opening Remarks

9:35 Managing Change Across Interdisciplinary Scientific Integration

Paul Konstant, Manager, Informatics Center of Excellence, Janssen Pharmaceutical Companies of Johnson & Johnson

Interdisciplinary scientific data integration efforts have faced numerous challenges as they attempt to keep pace with the evolution of both science and the technology that supports it. This talk will present a framework for identifying equivalence classes of protocols that reaches across disciplines based on easily understood scientific principles. Through the identification of the independent variables in a given experiment, patterns of parameter estimation, and the dependent variable outcomes, we will discuss how seemingly synonymous terminology. The principles of this framework allow well-structured expansion and graceful change management.

10:05 Integrating Clinical and Molecular Data: CASI Project

Brenda Yanak, Ph.D., Head, Translational Medicine Informatics & IT, Merck Research Labs

10:35 Coffee Break

11:00 Sponsored Presentation *(Opportunity Available)*

11:30 Visualization and Statistical Pharmacovigilance Methods Using Informatics for Integrating Biology and the Bedside (i2b2)

Shawn Murphy, M.D., Assistant Professor, Neurology, Harvard Medical School; Medical Director of Research Computing, Information Services, Partners Healthcare Inc.

The goal of i2b2 is to provide clinical investigators broadly with the software tools necessary to collect and manage hospital-based clinical research data in the genomics age as a cohesive entity—a software suite to construct and manage the modern clinical research chart. In the i2b2 software framework (the Hive), we are developing components (cells) to allow clinical investigators to compare medical outcomes on two matched groups of subjects. The components can be used for pharmacovigilance studies to detect associations between drugs and adverse events using Electronic Health Record data.

12:00 Translational Drug Discovery and Development: An Expedition of Data Integration and Utilization from Preclinical to Clinical

Dongzhou (Jeffery) Liu, Ph.D., Principal Clinical Investigator/Core Project Leader, Medical Affairs/Clinical Development, GlaxoSmithKline

This presentation will discuss contemporary reform and transformation in the drug industry, the latest strategies and approaches to renovate the process of drug R&D, and applications of combined approaches for efficient drug development.

12:30 Lunch for Purchase in the Exhibit Hall

13:45 Dedicated Poster Viewing in the Exhibit Hall

ONTOLOGIES AND STATE-OF-THE-ART WEB TECHNOLOGIES: BRING THE KNOWLEDGE TO THE USER

14:30 Chairperson's Remarks

14:35 Keeping Track of Data and What the Data are About: An Exploration in Realism-Based Ontology for Translational Research

Werner Ceusters, Ph.D., Director, Ontology Research Group, NYS Center of Excellence in Bioinformatics & Life Sciences

The Innovative Medicines Initiative (IMI) is a unique pan-European public and private sector collaboration between large and small biopharmaceutical and healthcare companies, regulators, academia and patients. One specific aim is to create a Knowledge Management Platform to arrive at effective data integration and analysis tools. We will discuss the exact place of ontology in such an endeavor and its relationship to terminologies and information models.

15:05 Sponsored Presentation *(Opportunity Available)*

15:35 Refreshment Break

16:00 Sponsored Presentation *(Opportunity Available)*

16:30 How Can a "Normal End User" Become a Web Service Consumer?

Christian Hauck, Ph.D., Knowledge Management and Competitive Intelligence, Novartis Pharma AG

Web services exist in various variants, but are mostly used by experts to build machine-to-machine interfaces. Without enough consumers, providers don't add value. Without enough providers, consumers get no value. We try to build web service consumer "frontends" for educated scientists and business people to help overcome this "chicken-and-egg" bootstrap problem. Attendees will learn about how to re-use web services that might already exist in their organizations, thereby creating more value out of what is already available.

17:00 Efficient Web-Browsing for the Life Scientist

Reinhard Schneider, Ph.D., Team Leader, Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL)

Anyone who regularly reads life science literature often comes across names of genes, proteins, or small molecules that they would like to know more about. We have developed a new service called Reflect (http://reflect.ws) that can be installed as a plug-in into browsers. Reflect tags gene, protein, and small molecule names in any web page, typically within a few seconds. Clicking on a tagged gene or protein name opens a popup showing a concise summary that includes synonyms, database identifiers, sequence, domains, 3D structure, interaction partners, subcellular location, and related literature. The popups also allow navigation to commonly used databases.

17:30 A Data Warehousing Approach for the Management of Clinical and Operational Information: Specific Challenges within Pharma Development

Norbert Fritz, Ph.D., Development Lead, Business Intelligence Warehouse, Information Management - Product Development Operations, F. Hoffmann-La Roche Ltd

The development of new medicines implies more and more extensive generation and usage of medical information. Management of this broad scope of information requires integration of data generated in heterogeneous data sources and presentation of clinical and operational data in one consolidated data model. Specific challenges for Pharma Development include: evolving and changing medical concepts, disparate systems generating source data, disparate data models and standards within various data sources, transient data quality issues, the need for frequent data refreshes. We will analyze the specific challenges, present approaches to design adequate solutions and discuss real-life experiences.

18:00 Close of Conference

SPONSORSHIP & EXHIBIT INFORMATION

Whether you are ready to present an exciting new technology, preparing for a new product launch, or need feedback on a specific idea, Bio-IT World Europe offers the perfect platform for you to present to a high-level, targeted audience.

The Biotechnica exhibit hall will host 13,000 attendees over the course of the event. Co-location with Biotechnica will allow you to exhibit as part of the larger event and reach your target audience in the Bio-IT World Europe session rooms, with an expected attendance of 300 delegates.

Exhibit in the Bio-IT World Pavilion in the Biotechnica hall, and you will be located in the central location for all Bio-IT World delegates. Traffic-building programs will be in place to ensure that delegates visit this pavilion.

Sponsors will get the opportunity to participate in three networking events offered to you free-of-charge by Biotechnica & Cambridge Healthtech Institute:

- Monday evening – pre-conference keynote presentation & reception
- Tuesday evening – Bio-IT World attendees have an exclusive dinner reception held at the convention center within close proximity to the session rooms
- Wednesday evening – a second social hosted by Biotechnica in the Bavarian Beer Hall, complete with dinner and a live band

These receptions are an excellent opportunity to network with your target audience. Attendance is included in selected sponsorship packages.

SPONSORSHIP OPPORTUNITIES:

Podium Presentations

A 15 or 30 minute podium presentation as part of the main conference.

(May also include a table-top in the foyer during the exclusive Bio-IT World Tuesday evening dinner reception.)

Coffee Breaks (exclusive per break)

Coffee breaks will be held in close proximity to the conference sessions. Table-tops will be available for sponsoring company to display corporate product literature.

Session Chair (exclusive per session)

An executive from your company will chair a session (a group of talks) on the main conference program. Includes a brief introduction to the entire session and the individual speakers.

EXHIBITOR INFORMATION

Exhibitors at Bio-IT World Europe will enjoy facilitated networking opportunities with more than 300 high-level decision-makers. Speak face-to-face with prospective clients and showcase your latest product, service or solution.

Marketing support from CHI and Biotechnica will include:

- Combined brochure mailings of 150,000
- E-mail campaigns of 1 million impressions

For more information on sponsorship and exhibit opportunities, please contact:

Katelin Fitzgerald
Manager, Business Development
+1 781-972-5458
kfitzgerald@healthtech.com

HOTEL & TRAVEL INFORMATION

Conference Venue:
Hannover Exhibition Grounds
Deutsche Messe
Messegelände
30521 Hannover
GERMANY

Please go to the following website for general visitor info and to make a hotel reservation:

http://www.biotechnica.de/visitorservice

SPONSOR DESCRIPTIONS

BIOTEAM
Enabling Science

www.bioteam.net

As a high-performance consulting practice, BioTeam delivers solutions to address the IT, storage, and data management challenges faced by scientists adopting next-generation sequencing technologies. We are dedicated to delivering objective, technology agnostic solutions to the life science researchers. From objective counsel to hands-on implementation, BioTeam enables science by bringing together technologies that turn research into results.

www.hp.com/go/lifesciences

For researchers working to find a cure, explore a theory, or discover a miracle, HP can help. We deliver technologies and solutions today that are accelerating the scientific breakthroughs of tomorrow. HP creates new possibilities for technology to have a meaningful impact on people, businesses, governments and society. As the world's largest technology company, HP brings together a portfolio that spans printing, personal computing, software, services and IT infrastructure to solve customer problems.

Your conference registration includes access to the Biotechnica Exhibition Halls!

ABOUT BIOTECHNICA

BIO TECHNICA

Geared to "Evolution of business and research," BIOTECHNICA 2010 invites you to Europe's leading gathering for biotechnology and life sciences, staged annually in Hannover, Germany. For three days the exhibition halls, the conference rooms and the partnering meeting boxes will be alive with exhibitors and visiting professionals from all over the world, together with investors and distinguished speakers from business, science and politics - all here to discuss the latest products, innovations, research findings and market opportunities.

Maximize your sales prospects. From research and product development, equipment, process technology and services to production and marketing: the exhibition section of BIOTECHNICA charts the biotech industry's value-adding chain from start to finish. Alongside the big industry players and SMEs, which have their own stands at the show, young and emerging biotech firms and scientific establishments are given ideal opportunities to showcase their work at the many group display stands representing national and international BioClusters and industry associations.

Reduce time and cost by making BIOTECHNICA the event of your choice in 2010. You can close deals, find business or research partners, discuss politics and forms of financing with experts and meet old and new friends - all this under one roof in just three days! All details at www.biotechnica.de.

HOW TO REGISTER: Online: Bio-ITWorldExpoEurope.com

Key Code 1076 F

Email: reg@healthtech.com • Phone: +1 781-972-5400 • Fax: +1 781-972-5425

1. REGISTRATION INFORMATION

❑ Mr. ❑ Ms. ❑ Mrs. ❑ Dr. ❑ Prof.

Name _____

Job Title _____ Div./Dept. _____

Company _____

Address _____

City/State/Postal Code _____

Country _____

Telephone _____

How would you prefer to receive notices from CHI? Email: ❑Yes ❑No Fax: ❑Yes ❑No

Email* _____ Fax _____

*Email is not a mandatory field. However, by excluding your email you will not receive notification about online access to pre-conference presenter materials, conference updates, networking opportunities and requested eNewsletters.

2. PRICING INFORMATION

SHORT COURSE PRICING

	Commercial	Acad., Gov., Hospital-affiliated	Student
1 Short Course	❑ €495	❑ €295	❑ €125
2 Short Courses	❑ €725	❑ €495	❑ €195

Monday, 4 October (10:00-13:30)

❑ Making Sense of Next-Gen Sequencing Data (10:00 – 13:30)

Monday, 4 October (Choose One) (14:00-17:00)

❑ Creating Synergy – Introduction to Biomedical Data Fusion (14:00 – 17:00)

❑ Cloud Computing for Life Sciences (14:00 – 17:00)

❑ Visualization of Large-Scale Biological Data (14:00 – 17:00)

EVENT PRICING

Early Registration Deadline until 16 July	❑ €1495	❑ €695	
Advance Registration Deadline until 3 September	❑ €1595	❑ €755	❑ €450
Registrations after 3 September and on-site	❑ €1745	❑ €805	

Please select the 2 conferences you're most likely to attend

5-6 October (Choose One)

❑ IT Infrastructure: Data Storage, Analysis and Visualization

❑ NGS Data Management: Sequencing Systems, Storage, and Analysis

6-7 October (Choose One)

❑ IT Infrastructure: Collaborative Software, Standards & Web Technologies

❑ Data Integration & Knowledge Management

INDIVIDUAL CONFERENCE PRICING

Early Registration Deadline until 16 July	❑ €995	❑ €495	
Advance Registration Deadline until 3 September	❑ €1095	❑ €545	❑ €300
Registrations after 3 September and on-site	❑ €1245	❑ €625	

Please select one conference:

5-6 October

❑ IT Infrastructure: Data Storage, Analysis and Visualization

❑ NGS Data Management: Sequencing Systems, Storage, and Analysis

6-7 October

❑ IT Infrastructure: Collaborative Software, Standards & Web Technologies

❑ Data Integration & Knowledge Management

COMPLIMENTARY BIOTECHNICA EVENTS

REQUIRED if you wish to attend these complimentary events. A ticket will be sent to you prior to the event. TICKETS NOT AVAILABLE ON-SITE.

❑ Monday, 4 October - BIOTECHNICA Opening and EUROPEAN BIOTECHNICA AWARD Ceremony plus Reception

❑ Wednesday, 6 October - BIOTECHNICA Night – Original Bavarian Beer Hall, full dinner reception, and band

DISCOUNTS

Poster Discount	❑ €35 off	❑ €35 off	❑ €35 off

REGISTER 3 - 4th IS FREE: Individuals must register for the same conference or conference combination and submit completed registration form together for discount to apply. Please reproduce this registration form as needed.

GROUP DISCOUNTS AVAILABLE! Special rates are available for multiple attendees from the same organization.

For more information on group discounts contact **David Cunningham at +1-781-972-5472**

❑ Please send me information on BIOTECHNICA's Partnering, an online networking tool

❑ I cannot attend but would like to purchase the Bio-IT World Europe Conference & Expo conference CD for €600 (plus shipping Massachusetts delivery will include sales tax.

3. PAYMENT INFORMATION

❑ Enclosed is a check or money order payable to Cambridge Healthtech Institute, drawn on a U.S. bank, in U.S. currency.

❑ Invoice me, but reserve my space with credit card information listed below.

Invoices unpaid two weeks prior to conference will be billed to credit card at full registration rate. Invoices must be paid in full and checks received by the deadline date to retain registration discount. If you plan to register on site, please check with CHI beforehand for space availability.

❑ Please charge: ❑ Visa (13-16 digits) ❑ MasterCard (16 digits)

Card # _____

Cardholder _____

Signature _____

Cardholder's Address (if different from above) _____

City/State/Postal Code/Country _____

CAMBRIDGE HEALTHTECH MEDIA GROUP

Yes! I would like to receive a FREE eNewsletter subscription to: **www.chimediagroup.com**

❑ **Weekly Update**
The latest industry news, commentary and highlights from Bio•IT World

❑ **eCliniqua**
Innovative management in clinical trials

Present a Poster and Save €35!

Cambridge Healthtech Institute encourages attendees to gain further exposure by presenting their work in the poster sessions.

To secure a poster board and inclusion in the conference materials, your abstract must be submitted, approved and your registration paid in full by **1:00 pm EDT, 8 September, 2010.** Register online, or by phone, fax or mail. Indicate that you would like to present a poster and you will receive abstract submission instructions via email.

I am interested in presenting a poster at
❑ **Bio-IT World Europe Conference & Expo**

Title _____

_____

CHI Insight Pharma Reports

A series of diverse reports designed to keep life science professionals informed of the salient trends in pharmaceutical technology, business, clinical development, and therapeutic disease markets. For a detailed list of reports, visit **InsightPharmaReports.com**, or contact Rose LaRaia, rlaraia@healthtech.com, +1-781-972-5444.

Barnett Educational Services

Barnett is a recognized leader in clinical education, training, and reference guides for life science professionals involved in the drug development process. For more information, visit **www.barnettinternational.com**.

Additional Registration Details

Each registration includes all conference sessions, posters and exhibits, food functions, and access to the conference proceedings link.

Handicapped Equal Access

In accordance with the ADA, Cambridge Healthtech Institute is pleased to arrange special accommodations for attendees with special needs. All requests for such assistance must be submitted in writing to CHI at least 30 days prior to the start of the meeting.

Substitution/Cancellation Policy

In the event that you need to cancel a registration, you may:

- Transfer your registration to a colleague within your organization. Credit your registration to another Cambridge Healthtech Institute program.
- Request a refund minus a €75 processing fee per conference.
- Request a refund minus the cost (€600) of ordering a copy of the CD .

NOTE: Cancellations will only be accepted up to two weeks prior to the conference.
Program and speakers are subject to change.

Video and or audio recording of any kind is prohibited onsite at all CHI events.

Cambridge Healthtech Institute
250 First Avenue, Suite 300, Needham, Massachusetts 02494
T: +1 781-972-5400 or toll-free in the U.S. 888-999-6288
F: +1 781-972-5425 • www.healthtech.com