

PRISME Forum

Pharmaceutical R&D Information Systems Management Executives

PRISME Forum TECH MEETING

PRISME Forum Chair: Matteo di Tommaso, *Pfizer*

May 14-15, 2014

North Mymms Park, UK

Host: GlaxoSmithKline

Download our PRISME Forum 2014 TECH Meeting App



or access it through: <http://my.yapp.us/PRISMETECH>

Registration

<http://www.surveygizmo.com/s3/1549765/PRISMEspring2014>

Meeting Venue

All sessions and networking events will be held at GSK's North Mymms Park conference center located Off Tollgate Road, Hatfield, Hertfordshire AL9 7TR.

Contacts

Program Coordinator: +44.77.68.173.518 / programme@prismeforum.org

Secretariat: office: +1.224.938.9523/mobile: +1.312.622.1234/logistics@prismeforum.org

PRISME Forum Tech Meeting Steering Committee

John Apathy, *VP, R&D Informatics, Celgene*

Dan Chapman, *Senior Principal Scientist, Software Development and Architecture, UCB*

Martin Leach, *VP, R&D Information Technology, Biogen Idec (Tech Meeting Steering Committee Chair)*

David Sedlock, *Senior Director, R&D Systems, Takeda Pharmaceuticals USA*

PRISME Forum Host

The Steering Committee of the PRISME Forum would like to thank GlaxoSmithKline, for hosting its spring 2014 meeting



PRISME Forum Statement of Compliance

“All meetings, working groups and communications will be open to all Members and any records thereof will be non-confidential and available for inspection by any Member. The Members acknowledge that discussing any commercially sensitive topics, including costs, volumes, inventories, sales level methods, channels of distribution, access to future products, markets, current or future prices, profitability, **contract pricing or trading terms** is prohibited. The Members of PRISME will strictly comply with all laws relevant to their activities, including US state and federal anti-trust laws and European competition laws.”

THEME:

Visualizing the Value of Big Data

“The great challenge in information visualization is to transform the abstract and intangible into something visible, concrete, and meaningful to the user.”[1]

The management and analysis of Big Data – or Big BioData – requires new and advanced computer-based technologies to cope with the volume of data, its variety and its velocity of change. New technologies to capture, store and search have been developed and are continually developing in this domain which is currently the center of much interest and excitement. Similarly analytical tools have been deployed and are under development. New and improved visualization approaches are required if insights and learnings from the Big Data environment are to be derived to the end-user community.

This PRISME Forum Technical Meeting will build on the knowledge developed at its Technical Meetings addressing “Big Data” (Spring 2013) and “Real World Evidence” (Fall 2013). The technical meeting will look at the opportunities to gain biopharma insight using BigData and visualization technologies. Such insight is essential to speed up the rate of innovation and reduce the costs of developing new and cost-effective therapies for unmet medical needs.

Analysis and visualization technologies are rapidly evolving to gain insight from large and complex data sets and the technologies are also evolving to provide more rapid real-time use of information. Furthermore, the open-source and academic community are very active in the Big Data paradigm.

The PRISME Forum has invited Martin Leach (VP R&D IT, Biogen Idec) to lead the Technical Meeting scientific advisory committee on which Dan Chapman (Sr. Principal Scientist, Software Development and Architecture, UCB), David Sedlock (Sr. Director, R&D Systems, Takeda Pharmaceuticals USA), and John Apathy (VP, R&D Informatics, Celgene) have all very kindly agreed to serve.

[1] Mapping Scientific Frontiers: The Quest for Knowledge Visualization. Chaomei Chen, Springer-Verlag January 2003

PROGRAM

WEDNESDAY, May 14, 2014

18:00	<i>Tour of North Myrms Park</i>
18:00	<i>Poster and Vendor Installation</i>
18:30	<i>Joint PRISME Forum Business Group + Tech Group Dinner</i>

THURSDAY, May 15, 2014

07:30	<i>Check-in and breakfast (Conservatory)</i>	
08:15	Welcome Notes	Matteo di Tommaso , VP, Research Business Technology, <i>Pfizer</i>
08:30	Introduction	Martin Leach , VP, R&D Information Technology, <i>Biogen Idec</i>
08:45	SESSION 1: PERSPECTIVES	CHAIR: John Apathy, VP, R&D Informatics, <i>Celgene</i>
08:45	The Emerging Role of Data Science in Pharma: How to Harness this Transformative Practice	Sarah Aerni , Senior Data Scientist, <i>Pivotal</i>
09:10	Visualization of Site Performance in Clinical Trials	Dimitris K. Agrafiotis , Vice President & Chief Data Officer, <i>Covance</i>
09:35	Big Data Imagery Analysis and Application to Life Sciences	Peter Guerra , Principal, and Rick Whitford , Chief Technologist, Data Science, <i>Booz Allen Hamilton</i>
10:00	Session 1 Panel Discussion	
10:20	<i>Coffee Break</i>	
10:50	SESSION 2: POSTER PRESENTATIONS	CHAIR: Craig Tulig, Director, Research Informatics, <i>Takeda Pharmaceuticals</i>
	P1: Data Viz: A Picture Says a 1000 Terabytes	Andrew Brosnan , Senior Analyst, Health Sciences, <i>Ovum</i>
	P2: Researcher-led Information Systems to Support Real Translational Workflows at the Cambridge Biomedical Campus	Peter Maccallum , Head of IT & Scientific Computing, <i>CRUK (Cambridge)</i>
	P3: Radar: A View of Complex Related Data	Frédéric Vanclef , Software Architect, Informatics, <i>UCB</i>
	P4: Integrating and Visualizing Big Data for Discovery at Takeda Pharmaceuticals	Jason Gratt , Associate Principal Software Engineer, <i>Takeda Pharmaceuticals</i>
	P5: Integrated Genomics Platform: Putting Patients and their Genomes into the Focus of our Research	Philip Groth , IT Business Partner, <i>Bayer HealthCare</i>
	P6: Getting a 'Heads-Up': EBI's High-Level Gene & Protein Summary Service	Jenny Cham , User Experience Analyst, <i>EMBL-EBI</i>
	P7: Extracting Knowledge from Images: Informatics Workflow for High-Throughput Microscopy	Shantanu Singh , Computational Biologist, <i>Broad Institute of Harvard and MIT</i>
	P8: Unbiased Disease Stratification within the IMI U-BIOPRED Severe Asthma Program Using Topological Data Mining	Dominic W. Burg , Post-Doctoral Fellow, <i>University of Southampton & The U-BIOPRED Consortium</i>
	P9: DRUG Process Design Improvement Based on DATA Management and Analysis	Eric Rulier , Senior Manager IT, Manufacturing & Labs Analytics, <i>UCB</i>
12:30	<i>Lunch</i>	
13:00	SESSION 3: VENDOR DEMONSTRATIONS	CHAIR: Martin Leach, VP, R&D Information Technology, <i>Biogen Idec</i> Each vendor will provide 15-minute demonstrations scheduled as follows: 13:00, 13:15, 13:30
	V1: "Here be dragons": Navigating the Waters of Big (Molecular Biology) Data Using Open Genomic Maps	Will Spooner , CTO, <i>Eagle Genomics</i>
	V2: Maximizing Value from Data: Making Better Decisions in a High-Dimensional World	Mark Roberts , Senior Analytics Consultant, <i>Tessella</i>
	V3: Insight and Visualization - Deliver Today Not Tomorrow/Getting to Insights in RWE Faster Using Visualization and Cloud Technology	Gary Hobbs , Sector Manager, <i>QlikView</i> , James Mucklow , Member of the Management Group, and David Berry , Managing Consultant, <i>PA Consulting</i>
	V4: Applications of Ayasdi Software to Precision Medicine, Drug Discovery and Biomarker Identification	Pek Lum , Chief Data Scientist, VP of Solutions, <i>Ayasdi</i>
	V5: A "Walled Garden" Collaborative Research Environment in the World of Translational Research	David Brown , CTO, Life Sciences, <i>BT Global Services</i>
13:45	SESSION 4: KEYNOTE PRESENTATION	CHAIR: Matteo di Tommaso, VP, Research Business Technology, <i>Pfizer</i>
13:45	Visualizing the Value of Big Data at EBI	Professor Dame Janet Thornton , FRS, Director, <i>EMBL-EBI</i>
14:30	SESSION 5: BRINGING IT ALL TOGETHER	CHAIR: Martin Leach, VP, R&D Information Technology, <i>Biogen Idec</i>
14:30	Round table discussions	
15:30	<i>Coffee Break</i>	
16:15	SESSION 5 (cont.)	CHAIR: Martin Leach, VP, R&D Information Technology, <i>Biogen Idec</i>
16:15	Read out of round table discussions	
16:45	SESSION 6	CHAIR: Dan Chapman, Senior Principal Scientist, Software Development and Architecture, <i>UCB</i>
16:45	Visualizing the Internet—the Akamai Approach	David Belson , Senior Director, Industry & Data Intelligence, <i>Akamai Technologies</i>
17:15	AWARDS AND RECEPTION	
17:15	<i>Awards</i>	
17:30	<i>Networking Reception (High Gallery)</i>	
19:00	Meeting Closes	

BIOS and ABSTRACTS

PRISME Forum Chair

Matteo di Tommaso, VP, Research Business Technology, *Pfizer*



Matteo leads Research Business Technology for Pfizer where he is responsible for strategy and implementation of IT and informatics services for Pfizer Research. At Pfizer, he has led efforts on cloud solutions for high performance computing (HPC), systems integration resulting from mergers and acquisitions, systems separations as a result of divestitures and IPOs, drug discovery data for decision making and insight, translational informatics solutions for patient stratification and integration of clinical and molecular data, laboratory automation services, and data center simplification. His efforts in pre-competitive collaboration have led to opensource tools for Chemistry eNotebook and biomolecule discovery and contributions to efforts including OpenBEL, Pistoia Alliance and tranSMART.

Before joining Pfizer, in 2004, Matteo led the team at Celera Genomics responsible for building Celera's scientific information products and Applied Biosystems' eCommerce solutions. Prior to that he led the development of the "SeqStore" product line for Genetics Computer Group, a set of products and services for pharmaceutical drug discovery. GCG, he spent 3 years at the European Bioinformatics Institute (EBI) at the start of the institute in Cambridge, UK. Matteo began his career in IT, at Warner-Lambert Parke-Davis, with a degree in Chemistry from Indiana University, leading the migration and replacement of pre-clinical information systems to improve data quality and usability.

Tech Meeting Steering Committee Chair

Martin Leach, VP, R&D Information Technology, *Biogen Idec*



Martin Leach's role at Biogen Idec includes support of IT for Basic Research/Discovery through and including support of Clinical Trials, Safety, Regulatory and Global Medical Affairs. He is responsible for building out a new Data Sciences Group that will initially support the needs of Research but will expand over time to leverage these capabilities to support all areas across the enterprise. Working jointly with the SVP of Translational Research, Martin will have dotted line responsibility for Computational Biology. Prior to Biogen Idec, Martin led the administrative IT and computational infrastructure at The Broad Institute of MIT and Harvard, based in Cambridge, MA (responsible for more than 13PB storage and several large compute farms totaling over 10,000 cores).

Prior to The Broad Institute, he led IT in support of the research environment for Discovery & Preclinical Sciences at Merck. Prior to Merck, his work at Booz Allen was working with the partnership to establish and build out the PharmaIT practice. This spanned the entire pharma value chain with projects such as post-merger integration, IT strategy, informatics strategy, organizational change and design.

Martin has Ph.D. In Molecular Pharmacology of Neurotransmitter Receptors from Boston University School of Medicine and a B.Sc. (Hons – 1st class) in Cellular & Molecular Sciences from Cambridgeshire College of Arts & Technology (renamed Anglia Ruskin University – Cambridge).

SESSION I: PERSPECTIVES

Chair: John Apathy, VP, R&D Informatics, *Celgene*



John joined Celgene in 2013, leading its Global R&D Informatics efforts for the IT organization. He brings over 25 years of experience in Pharmaceutical new product and R&D capability development across the Pharmaceutical/Biotech, Life Sciences and Management Consulting industries. John has extensive knowledge and experience based upon leadership positions within both industry and consulting at Celgene, GlaxoSmithKline, Pfizer, Wyeth, Accenture, Eli Lilly & Company, PA Consulting Group, and Solutia Pharmaceutical Advisors. He is also an experienced leader in strategy, processes, and systems for development and commercialization of new pharmaceutical products—skills and experience include delivery of strategy, large-scale program management, product development processes, life-cycle extension, product launch (particular emphasis on the Early Development phase of Pharmaceutical R&D, combining technical knowledge with change leadership, consulting problem-solving, strategy development, and change management processes).

Vendor Perspective:

The Emerging Role of Data Science in Pharma: How to Harness this Transformative Practice

Sarah Aerni, Senior Data Scientist, *Pivotal*

CRO Perspective:

Visualization of Site Performance in Clinical Trials

Dimitris K. Agrafiotis, Vice President & Chief Data Officer, *Covance*

Consultant Perspective:

Big Data Imagery Analysis and Application to Life Sciences

Peter Guerra, Principal, and Rick Whitford, Chief Technologist, Data Science, *Booz Allen Hamilton*

Sarah Aerni, Senior Data Scientist, *Pivotal*



Sarah Aerni is a member of the Pivotal Data Science team with a focus on healthcare and life science. She has a background in the field of Bioinformatics, developing tools to help biomedical researchers understand their data. She holds a BS In Biology with a specialization in Bioinformatics and minor in French Literature from UCSD, and an MS and PhD in Biomedical Informatics from Stanford University. During her time as a researcher she focused on the interface between machine learning and biology, building computational models enabling research for a broad range of fields in biomedicine. She also co-founded a start-up providing informatics services to researchers and small companies. At Pivotal she works with customers in life science and healthcare building models to derive insight and business value from their data.

Vendor Perspective:

The Emerging Role of Data Science in Pharma: How to Harness this Transformative Practice

Across all industries, big data is associated with innovation and progress. In the healthcare and life science sector, the concept of big data refers not only to the volume of the data available, but also its diversity. Emerging technologies in this space are able to store and process large volumes of data, but the key to advancement in the understanding, management and treatment of disease requires bringing them together and learning to make sense of them, together. Creating successful models and insights from this data requires bringing practitioners, such as data scientists, together with the clinical researchers and other domain-specific communities, the data and the tools that enable them to work most efficiently and effectively, to collectively power meaningful innovation.

Pivotal focuses squarely on this challenge, of arming the data scientists with the tools and technologies to solve problems ranging from analyzing large-scale clinical data to predict patient outcomes. Using faster technology, the cost of asking questions of the data is reduced, and with the correct tools, modeling and integrating the data at scale becomes possible. It is with these tools and our process of discovery, feature engineering and modeling, that we have solved problems from predicting patient length of stay, to understanding how to alter the potency of a pharmaceutical product using manufacturing data.

In this talk we will share our experiences working with the process, tools and technologies Pivotal develops to enable discovery and predictive modeling in the healthcare and life science space.

Dimitris K. Agrafiotis, Vice President & Chief Data Officer, Covance



Dimitris K. Agrafiotis, PhD, FRSC, is Vice President & Chief Data Officer at Covance, where he is responsible for the architecture, design, implementation, operation, and support of information technology services, and for establishing an encompassing informatics vision to infuse innovation and strategic differentiation into the company's product offerings.

Prior to joining Covance in 2013, he spent 10 years at Johnson & Johnson Pharmaceutical Research & Development where he served as Vice President of Informatics and Research IT. Dr. Agrafiotis received his PhD in Theoretical Organic Chemistry from Imperial College, London (1988, Henry Rzepa advisor), and held postdoctoral fellowships with Andrew Streitwieser at the University of California, Berkeley, and Nobel laureate E. J. Corey at Harvard. In 1991, he joined Parke-Davis as a Senior Scientist in the CADD group, and in 1994 he moved to 3-Dimensional Pharmaceuticals (3DP) as a founding member of its scientific staff, responsible for building the company's informatics and computational drug design capabilities. Following a successful IPO and the acquisition of 3DP by Johnson & Johnson in 2003, he was appointed Senior Research Fellow and Team Leader of the Molecular Design & Informatics group at the East Coast site, a position he held until 2006 when he was appointed global head of Informatics and Research IT. His work spans many areas of computational life sciences and informatics, and is documented in over 90 peer-reviewed publications and book chapters, and 18 issued US patents.

He frequently delivers invited lectures, organizes national and international symposia, and serves on the editorial boards of several prominent journals in his field. He is a member of the Covance Science Council and serves as Adjunct Professor of Informatics at Indiana University School of Informatics and the University of Athens School of Pharmacy. In 2012 he was elected Fellow of the Royal Society of Chemistry in recognition for his contributions to chemical research.

CRO Perspective:

Visualization of Site Performance in Clinical Trials

As a market leader in central laboratory and pre-clinical services and top 5 provider of phase III clinical trial management services, Covance has assembled the most comprehensive investigator knowledge-base in the pharmaceutical industry, spanning >11,000 protocols, >600 indications, >175,000 unique investigators, and >14 million patient visits over the past 10 years. Recognizing the power of data-driven decision making, we are assembling a world-class informatics capability to unleash the enormous potential of this data to help pharmaceutical and biotechnology companies bring the miracles of medicine to the people in need faster. In this talk, we describe how we utilize information from our clinical labs to optimize the selection of clinical sites and investigators based on historical performance data in the disease areas and clinical indications of interest. Our approach addresses questions such as whether it is feasible to recruit patients to the current study design, how to optimize site networks based on historical performance data and available patient populations, how to be more precise and confident about country allocation, how to optimize the inclusion and exclusion criteria, how to predict enrollment based on epidemiology and prevalence at a specific location, and how to identify and develop the next generation of top-tier investigators.

Peter Guerra, Principal, *Booz Allen Hamilton*



Peter Guerra is one of the key leaders of Booz Allen's Data Science practice. He specializes in the design and development of highly available, large-scale distributed systems and advanced analytics, and is responsible for leading several large-scale analytics and big data computing projects. He works with clients across wide range of industries including National Intelligence, Defense, Commercial Health and Financial Services. Mr. Guerra is currently completing his MBA at Loyola University, and possesses a B.A. degree in English and B.S. degree in Computer and Information Science both from University of Maryland.

Rick Whitford, Chief Technologist, Data Science, *Booz Allen Hamilton*



Rick Whitford is seasoned technology professional with expertise in software development, systems and network engineering and enterprise architecture and design spanning several industries. He most recently led the design, development, and implementation of a large scale image analysis platform for several US Federal clients. He has also focused on transitioning legacy systems to Big Data/ NOSQL architectures, enabling more complex queries, providing for advanced analytics and allowing for data driven business decisions. Mr. Whitford currently holds a MS in Information Technology with a concentration in Information Assurance from the University of Maryland, and an MBA from the Robert H. Smith School of Business.

Consultant Perspective:

Big Data Imagery Analysis and Application to Life Sciences

Over the last 5 years Booz Allen has been working with numerous US Federal and Commercial clients to develop advanced imagery analytics that are capable of processing millions of images for automated analysis. These systems were designed to perform feature analysis across petabytes of disparate images and, when linked with other information (text, network graphs, and so forth) and combined with advanced visualizations, allows researchers to quickly achieve a high degree of automated analysis. We have used these same architectures and design patterns to build solutions within Life Sciences to address challenges in data integration and search/retrieval. This talk will highlight the technical aspects of the system, talk through the challenges of analyzing large imagery data sets, and the interesting applications big data architectures and advanced visualizations/ user experiences have within Life Sciences based on case studies.

SESSION II: POSTER PRESENTATIONS

Chair: Craig Tulig, Director, Research Informatics, *Takeda Pharmaceuticals*



Craig started in Life Sciences during the early days of genomics, where he build the bioinformatics platform for a small biotech startup, which was in the business of sequencing pathogens and doing human genetic research by positional cloning. He later joined the Cheminformatics group of Millennium Pharmaceuticals where he led the development of the company's main decision support system for Drug Discovery. He came to Takeda through its acquisition of Millennium where he is now the Director of Research Systems.

- P1: Data Viz: A Picture Says a 1000 Terabytes**
Andrew Brosnan, Senior Analyst, Health Sciences, *Ovum*

- P2: Researcher-led Information Systems to Support Real Translational Workflows at the Cambridge Biomedical Campus**
Peter Maccallum, Head of IT & Scientific Computing, *CRUK (Cambridge)*

- P3: Radar: A View of Complex Related Data**
Frédéric Vanclef, Software Architect, Informatics, *UCB*

- P4: Integrating and Visualizing Big Data for Discovery at Takeda Pharmaceuticals**
Jason Gratt, Associate Principal Software Engineer, *Takeda Pharmaceuticals*

- P5: Integrated Genomics Platform: Putting Patients and their Genomes into the Focus of our Research**
Philip Groth, IT Business Partner, *Bayer HealthCare*

- P6: Getting a 'Heads-Up': EBI's High-Level Gene & Protein Summary Service**
Jenny Cham, User Experience Analyst, *EMBL-EBI*

- P7: Extracting Knowledge from Images: Informatics Workflow for High-Throughput Microscopy**
Shantanu Singh, Computational Biologist, *Broad Institute of Harvard and MIT*

- P8: Unbiased Disease Stratification within the IMI U-BIOPRED Severe Asthma Program Using Topological Data Mining**
Dominic W. Burg, Post-Doctoral Fellow, *University of Southampton & Tsdhe U-BIOPRED Consortium*

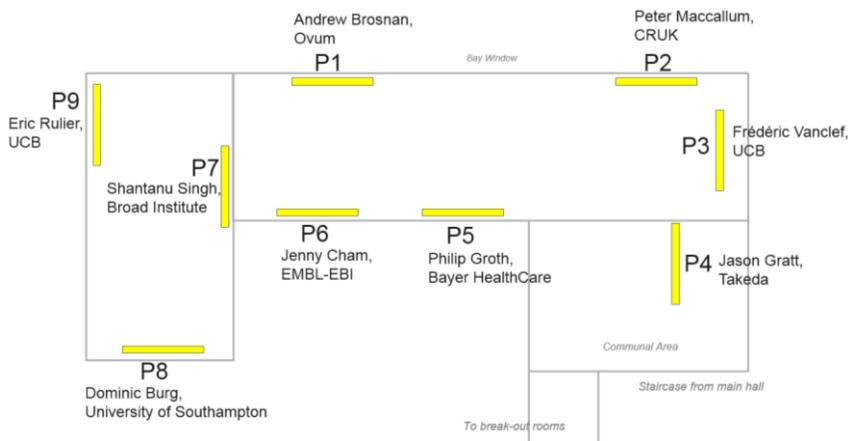
- P9: DRUG Process Design Improvement Based on DATA Management and Analysis**
Eric Rulier, Senior Manager IT, Manufacturing & Labs Analytics, *UCB*

SESSION STRUCTURE

At 10:50 am, Session Chair will invite participants to take their seats.

Chair will open the session by setting the poster presentation within the larger theme of the event and will present the structure and flow.

Chair and PRISME Forum staff will ring the bell on each rotation (minute 8 of each presentation and then minute 10 at which time the presentation must end).



Delegates are invited to identify the color of their badge and match to rotation called out loud by staff: "Rotation 1, 2, ...n". Rotations will involve shift of participants' groups to the next poster on their right.

ROTATIONS:

ROTATION 1 - 11:00

P1 - Turquoise
P2 - Orange
P3 - Green
P4 - Yellow
P5 - Black
P6 - Red
P7 - Purple
P8 - Brown
P9 - White

ROTATION 4 - 11:30

P1 - Purple
P2 - Brown
P3 - White
P4 - Turquoise
P5 - Orange
P6 - Green
P7 - Yellow
P8 - Black
P9 - Red

ROTATION 7 - 12:00

P1 - Yellow
P2 - Black
P3 - Red
P4 - Purple
P5 - Brown
P6 - White
P7 - Turquoise
P8 - Orange
P9 - Green

ROTATION 2 - 11:10

P1 - White
P2 - Turquoise
P3 - Orange
P4 - Green
P5 - Yellow
P6 - Black
P7 - Red
P8 - Purple
P9 - Brown

ROTATION 5 - 11:40

P1 - Red
P2 - Purple
P3 - Brown
P4 - White
P5 - Turquoise
P6 - Orange
P7 - Green
P8 - Yellow
P9 - Black

ROTATION 8 - 12:10

P1 - Green
P2 - Yellow
P3 - Black
P4 - Red
P5 - Purple
P6 - Brown
P7 - White
P8 - Turquoise
P9 - Orange

ROTATION 3 - 11:20

P1 - Brown
P2 - White
P3 - Turquoise
P4 - Orange
P5 - Green
P6 - Yellow
P7 - Black
P8 - Red
P9 - Purple

ROTATION 6 - 11:50

P1 - Black
P2 - Red
P3 - Purple
P4 - Brown
P5 - White
P6 - Turquoise
P7 - Orange
P8 - Green
P9 - Yellow

ROTATION 9 - 12:20

P1 - Orange
P2 - Green
P3 - Yellow
P4 - Black
P5 - Red
P6 - Purple
P7 - Brown
P8 - White
P9 - Turquoise

Andrew Brosnan, Senior Analyst, Health Sciences, *Ovum*



Andrew Brosnan is a senior analyst in the Ovum Health Sciences team. His research focuses on the use of technology in the pharmaceutical and biotechnology markets. Andrew has considerable experience working with life sciences applications. He has held a variety of positions and led several divisional and enterprise-wide IT projects. His experience includes working closely with scientific stakeholders as well as bench researchers, giving him a unique perspective on the technology needs and experiences of end users in a laboratory setting.

Prior to joining Ovum, Andrew was a project leader and application analyst in Merck and Co.'s drug discovery IT division. He was a subject matter expert for several applications within the drug discovery portfolio including data mining, data archival, and lab automation applications. He also started his own firm in which he provided IT consulting services to the pharmaceutical sector.

Andrew has additional experience in the financial sector in both a financial and IT capacity, having worked as a bank examiner for the US Treasury Department and holding several supervisory positions within financial IT departments. He holds a BA from Boston University and is currently enrolled in Warwick University's Executive MBA program.

P1: Data Viz: A Picture Says a 1000 Terabytes

To stay competitive in a fast paced and ever-changing business environment, pharmaceutical companies are hard pressed to make decisions at a faster pace, with larger and more complex datasets. Ovum believes to be successful in such an endeavor organizations will not just need to embrace new infrastructures and technologies to handle Big Data needs, but they will also aim to democratize the analytic environment to allow business user to easily gain insight from data, quicker. However, such agility is easier preached than achieved, leaving organizations stuck between a rock and a hard place: they need to work with more big and diverse data than before, while making the analysis as user friendly and self-serviced as possible. This dual requirement is generating demand shifts from static tables, graphs, and charts towards analysis methods that engage the end-user with interactive and agile visualizations.

During this poster session, Ovum's will highlight current trends in the business intelligence and analytic market relating to data visualization, data discovery, and the changing role of IT.

Peter Maccallum, Head of IT & Scientific Computing, *CRUK (Cambridge)*



Peter Maccallum is Head of IT & Scientific Computing in Cancer Research UK. He has a BSc in Chemistry from the University of Edinburgh and PhD in Biochemistry from the University of Glasgow on the computational analysis of protein structure. He has worked at Edinburgh Parallel Computing Centre and the European Bioinformatics Institute, and as a consultant and software architect in the life sciences industry for genomics, bioinformatics and pharmaceutical companies specialising in distributed systems applied to bioinformatics. Since 2008 he has been responsible for IT provision at Cancer Research UK.

P2: Researcher-led Information Systems to Support Real Translational Workflows at the Cambridge Biomedical Campus

The Cancer Research UK Cambridge Institute was founded to focus on translational impact, bridging the gap between basic biology and clinical application in cancer diagnosis and treatment. This requires the coordination of data collection and analysis between multiple clinical and laboratory environments, in multiple organizations. The multi-organizational nature of the work means a common informatics platform is not an option; the dynamic nature of the data and clinical confidentiality mean data warehousing is not a solution. On top of this, software developer time is in short supply.

Based on our experience with model-driven development approaches and the real-world issues in deploying and using systems, we have developed an approach based around models of data collection workflows. Researchers are trained to identify the points where entities are registered and relations created. The models focus on the key metadata – identifiers, relations, and where they are represented in tables. They can be straightforwardly represented as graphs for explanation, query construction or gap analysis.

Since implementation and maintenance of software is a significant barrier and cost, we have developed a process which allows researchers to build models using tools such as Excel and Access, and then lets IT staff quickly extend these to multi-tier systems using open source toolkits and virtualization. The resulting process efficiently puts domain modelling in the hands of the researchers, improves support for research systems, and ultimately improves the quality of data and the research derived from it.

Frédéric Vanclef, Software Architect, Informatics, *UCB*



Frédéric Vanclef is currently a Software Architect in Informatics at UCB. During his time at UCB, he has held many positions such as Project Manager, Application Development Manager, Associate Director Software Architecture, and Director of IT in Service Management and Process Improvement. Prior to joining UCB in 1997, Frédéric was a consultant in the bank sector. Frédéric has a degree in Chemical Engineering from UCL (Université Catholique de Louvain) in addition to a degree in Management.

Radar: A View of Complex Related Data

Highly related data is quite difficult to represent in two dimensions charts. Our challenge was presenting at a glance the relationships between diseases, targets, pathways, and compounds which have been identified during a drug repurposing exercise.

Our visualization offers the following features:

- Allowing the user to pick target after target;
- Appealing graph layout with clear and intuitive navigation flow;
- Clear depiction of contextual interconnection between entities;

After some significant time investments made in the generation of drug repurposing data, we decided to optimize the use of those data creating custom visualizations in the shape of a radar screen. This is instinctively well perceived by the user because it evokes the search and allows the representation of relative proximity of multiple variables to a center. The visualization was developed as a web page, using of the D3 SVG library (Data-Driven Document). The representation proved to be very efficient in showing relevant connected information upon user's selection and allowing backward selection for network expansion.

Jason Gratt, Associate Principal Software Engineer, *Takeda Pharmaceuticals*



Jason Gratt is an Associate Principal Software Engineer at Takeda (formerly Millennium) Pharmaceuticals. After earning his PhD in Polymer Chemistry at the Massachusetts Institute of Technology (where he also completed coursework in computer science and software engineering), he combined his interests and has been working for the past 15 years as a software architect in bioinformatics and cheminformatics, with a particular emphasis on big data, visualization technology, and decision support systems.

Integrating and Visualizing Big Data for Discovery at Takeda Pharmaceuticals

We present an overview of Takeda Boston's Discovery Informatics environment. The Discovery Data Warehouse integrates a variety of low and high volume data sources, including chemical registration (structures and electronic lab notebook links), drug source (purity and inventory), in silico calculated properties, in vitro assays (pending requests, individual results, and averages), in vivo data (PK/PD parameter measurements and report links), and patent info. The Discovery Data Viewer provides a fluid, easy to use visualization interface, with powerful filtering and drill-down capabilities, list views and single compound views, data export to Excel and SD formats, and specialized modules for selectivity analysis, SAR, and cluster analysis. In-memory server-side caching and AJAX web technologies enable fast performance, and embedded Spotfire templates (web player mashups) facilitate exploratory analysis via construction of scatter plots with dynamic selection of fields for X and Y axes and coloring. PharmaSee, a separate web application that is also powered by the Discovery Data Warehouse and makes use of embedded Spotfire templates, provides a similar interface for visualizing drug combination studies and constructing heatmaps of synergistic and antagonistic interactions in various cell lines. Finally, the Spotfire client is directly leveraged by end users to publish analyses of genomics and gene expression data from our NexGenomics database, for company-wide dissemination via the Spotfire web player.

Philip Groth, IT Business Partner, *Bayer HealthCare*



Dr. Philip Groth has a Master's degree in bioinformatics and a PhD in computer science from the Humboldt University Berlin. After working on yield prediction in genetically modified plants for two years he assumed a position as research scientist in the Oncology department of Bayer Pharma AG in Berlin. In 2013 he assumed a position as IT Business Partner for Oncology and Genomics at Bayer Pharma AG. He lives in Berlin, Germany with his wife and two children and in his spare time enjoys sailing and horseback riding.

Integrated Genomics Platform: Putting Patients and their Genomes into the Focus of our Research

The fast progress in the generation of genomic data has reached the patient. Especially the advent of next generation sequencing and high resolution microarrays enable accurate descriptions of diseases with a strong genetic component ultimately leading to novel therapeutic approaches. Application of these technologies, however, lead to large amounts of data in need of effective storage and analysis. As now several data types (mutation, expression, microRNAs) become available for each patient, patient-centric views and analyses become mandatory. Consistent data handling and storage is a scientific and technological challenge towards both the research organization and the IT infrastructure.

We are therefore establishing the Integrated Genomics Platform (IGP) as a strategic initiative to develop a new central platform for research genomics data from Cardiology, Oncology and Clinical Sciences. This platform supports advanced data analysis capabilities and is intended to simplify discovery processes, e.g. for novel therapeutic targets and genetic biomarkers.

We aim to overcome current bottlenecks to enable true translational research by establishing a global mandatory repository and toolbox for storage and analysis of genomics data as well as common standards for data annotation, privacy & security.

Jenny Cham, User Experience Analyst, *EMBL-EBI*



Jenny is based in the Web Development Team at EMBL-EBI. She is interested in applying user experience design practices to support the development of user-friendly services for life scientists. Jenny, originally a biochemist, has worked in pharmaceutical R&D at Merck KGaA (Darmstadt) and GSK (Stevenage) and has a doctorate in proteomic bioinformatics from Cranfield University. She is a regular speaker at user experience and bioinformatics meetings. (Blog: <http://jennycham.co.uk/>)

Getting a 'Heads-Up': EBI's High-Level Gene & Protein Summary Service

The EBI has now made it easier for life scientists to search and browse vast stores of publicly available biological data and related information. Its unique, intuitive search service provides a huge simplification for users exploring the data. With more than 300 million entries indexed and updated daily, the search provides an efficient gateway to explore major EBI data collections.

To develop the service, we carried out extensive user experience analysis, involving prototyping and usability testing with four pharmaceutical companies and seventeen academic institutes.

Shantanu Singh, Computational Biologist, *Broad Institute of Harvard and MIT*



Shantanu Singh, PhD, is a Computational Biologist at the Broad Institute of MIT and Harvard. He received his doctoral degree in computer science from the Ohio State University, where he used imaging, machine learning and visualization techniques to solve problems in cancer biology. He completed his post-doctoral training under Anne Carpenter at the Broad Institute where he developed computational tools for image-based profiling of genes and compounds - a line of research that he continues to build upon in his current role as a member of the Imaging Platform.

Extracting Knowledge from Images: Informatics Workflow for High-Throughput Microscopy

Microscopy images contain rich information about the state of cells, tissues, and organisms and are an important part of experiments to address a multitude of basic biological questions. Our laboratory works with dozens of collaborators around the world to design and execute image-based experiments, primarily high-throughput screens. These experiments test thousands of chemical or genetic perturbations in order to identify the causes and potential cures of disease.

In this presentation, I will discuss the informatics workflow for such projects. First, using the example of CellProfiler, an open-source image-analysis software developed by our group, I will outline a typical image-analysis pipeline that can be used to extract hundreds of measurements per cell from these large experiments. Next, I will describe how such large datasets are analyzed using standard statistical analysis for simple experiments, and machine learning algorithms for more complex experiments. Overall, I will focus on principles about the types of analyses are needed for image data, along with concrete examples and case studies.

Dominic W. Burg, Post-Doctoral Fellow, *University of Southampton & the U-BIOPRED Consortium*



Dominic Burg is a research fellow working in the IMI-UBIOPRED Consortium. Dominic received his PhD from the University of New South Wales, Sydney, and continued his post-doctoral training at the University of Sydney and Australian Proteome Analysis Facility. Dominic is currently based within the Centre for Proteomic Research, University of Southampton as an experimental and data informatics scientist for the proteomics program of U-BIOPRED. Dominic's research activities are wide ranging, but his particular interests are the interfaces between experimental, biological, and informatic sciences.

Unbiased Disease Stratification within the IMI U-BIOPRED Severe Asthma Program Using Topological Data Mining

Dominic Burg^{1,2}, Doroteya Staykova¹, Pek Lum³, Xian Yang⁴, Yike Guo⁴, Anthony Rowe⁵, Ratko Djukanović², Paul Skipp¹ and the U-BIOPRED consortium

¹. *Centre for Proteomic Research, University of Southampton, UK.*

². *NIHR Southampton Respiratory Biomedical Research Unit, Southampton University Hospital, UK.*

³. *Ayasdi Inc, USA.*

⁴. *Imperial College, London, UK.*

⁵. *Janssen R&D, UK*

The U-BIOPRED consortium is an EU-wide collective of academics, pharmaceutical companies and patient organisations working collaboratively to improve understanding of severe asthma. Representing the largest study cohort recruited for this disease, a variety of sample types are currently being analysed in parallel using a range of 'omics technologies to map molecular and clinical phenotypes of severe asthma. The heterogeneity of the disease, combined with the complexity of the study cohort (e.g. participants on a combination of medications, and varying co-morbidities), the range of biofluids and tissues analysed (each with corresponding challenges), and the multiple analytes being measured (e.g. lipids, proteins, mRNA) have necessitated the development of multiple data analysis pipelines to mine these complex datasets. One of the approaches used by the consortium is Topological Data Analysis (TDA), implemented via the Ayasdi software platform. TDA generates topological networks that allow the scientist to explore, condense, visualise and extract useful information from these complex and multi-modal data. Additionally, we have extended this approach as a method for unbiased feature selection prior to application of other machine learning techniques such as support vector machine for disease sub-classification.

We present the results of preliminary analysis of an exploratory proteomic dataset using the Ayasdi TDA software platform. Data were generated from induced-sputum and blood serum samples using high-resolution mass spectrometry (LC-IMS-MSE) and resultant spectra were processed and searched against a protein database to identify and quantify proteins. Additionally, processed 4D ion spectra (DRAM) were aligned and grouped across samples/participants using an in-house developed algorithm, allowing the analysis of spectral ion data (> 500,000 spectral ion features per experiment) independent of traditional protein database searching approaches. Both of these processed datasets were subsequently uploaded into the Ayasdi analytical platform, along with clinical and patho-biological information, for TDA. Groupings of participants from TDA were selected based on persistence of structure aided by pathobiological meta-data. These topological groupings were subsequently used as classifiers in other machine learning approaches, implemented in InforSense, and resulted in an improvement in predictive performance over cohort information alone.

Eric Rulier, Senior Manager IT, Manufacturing & Labs Analytics, UCB



Eric Rulier is Senior Manager IT, Manufacturing & Labs Analytics at UCB since 2011. Prior to UCB, Eric gained extensive experience in information laboratory systems (LIMS, ELN, CDS, labs standalone system) during the 20 years spent within GSK as an IT project leader as well as an Information System Business Team Leader in the pre-clinical department.

DRUG Process Design Improvement Based on DATA Management and Analysis

Most of scientific process data generated are not free of access, even if managed in databases. At UCB, data recently get available including its context. It allows process development scientist to draw easily design space and define critical parameters. To support investigation study as impact analysis, manufacturing dashboards and trends are automatically published. An example of correlation between process data and patients clinical responses will be presented as an illustration of advanced data analysis.

SESSION III: VENDOR DEMONSTRATIONS

Chair: Martin Leach, VP, R&D Information Technology, *Biogen Idec*

V1: “Here be dragons”: Navigating the Waters of Big (Molecular Biology) Data Using Open Genomic Maps

Will Spooner, CTO, *Eagle Genomics*

V2: Maximizing Value from Data: Making Better Decisions in a High-Dimensional World

Mark Roberts, Senior Analytics Consultant, *Tessella*

V3: Insight and Visualization - Deliver Today Not Tomorrow

Gary Hobbs, Sector Manager, *QlikView*

Getting to Insights in RWE Faster Using Visualization and Cloud Technology

James Mucklow, Member of the Management Group *and*

David Berry, Managing Consultant, Business Intelligence, *PA Consulting*

V4: Applications of Ayasdi Software to Precision Medicine, Drug Discovery and Biomarker Identification

Pek Lum, Chief Data Scientist, VP of Solutions, *Ayasdi*

V5: How You Can Provide a ‘Walled Garden’ Collaborative Research Environment in the World of Translational Research

David Brown, CTO, Life Sciences, *BT Global Services*

Will Spooner, CTO, *Eagle Genomics*



William Spooner is a seasoned operational bioinformatician with a track record of delivering tools for high-throughput genomics research. Having worked previously on Ensembl, BioMart, Gramene and WormBase at the European Bioinformatics Institute, the Wellcome Trust Sanger Institute and Cold Spring Harbor, his current focus is on making life easier for users of open source/data in commercial settings. His strategic thinking is driven by the huge opportunities for data analysis in the life sciences provided by the near simultaneous arrival of NGS and cloud computing.

“Here Be Dragons”: Navigating the Waters of Big (Molecular Biology) Data Using Open Genomic Maps

Genomics epitomizes big data. For the past decade the main visualization tool for molecular biologists has been the Genome Browser. The market is led by open source Browsers such as Ensembl whose success is in part attributable to a vibrant developer community which includes local bioinformatics specialist Eagle. Right now several of our Pharma partners are struggling with the same problem; the canonical reference sequence for a given gene is not always representative. You might, for example, be blindly optimizing your chemistry against the wrong cDNA/protein! The data needed to reconstruct the "correct" sequences for a given population are there in the database, so why not visualize them? This is exactly what we did, and this demo presents our tool for displaying population-specific cDNAs and proteins in Ensembl. Furthermore, the code has been cleared for release and will soon be contributed to Ensembl for the good of the wider community.

Mark Roberts, Senior Analytics Consultant, *Tessella*



problems.

Mark Roberts is an analytics consultant for Tessella in the UK. Mark has spent nearly a decade working with Tessella's pharmaceutical customers as a scientific computing consultant, business analyst and project manager, helping them to extract maximum value from their data and optimizing their workflows. This expertise and experience is being put to good use as he focuses on working with customers across multiple sectors to guide them through the emerging fields of big data and analytics. Before joining Tessella, Mark gained a BSc and PhD in Artificial Intelligence and Computer Science, specializing in the use of advanced machine learning techniques for solving optimization and image analysis

Maximizing Value from Data: Making Better Decisions in a High-Dimensional World

Modern high-end R&D produces vast quantities of complex data. Using that data to make useful decisions relies on the difficult tasks of being able to reduce its complexity, analyze it effectively and present it simply and appropriately. The recent surge in Big Data technologies has amplified these difficulties and has left many companies struggling to make good use of the data assets now at their disposal. How do you sensibly work with 1,000 dimensional data? This talk will discuss the issues involved in making best use of complex multi-parametric data, from analyzing it using mathematical or machine learning techniques, making good decisions under uncertainty using interactive causal network diagrams, projecting high-dimensional spaces into easily understandable visualizations, and avoiding common statistical pitfalls in the analysis and presentation of large datasets. The talk will be illustrated using examples from Tessella's diverse experience across different R&D sectors, delivering the right data, presented in the right way, to the right person, in order to make better business decisions.

Gary Hobbs, Sector Manager, QlikView



Gary has worked in the field of Business Intelligence and now Data Discovery for over 15 years working for the leading companies of the time such as BusinessObjects, SAP and for the last for years at QlikView.

Gary has worked with many of the world's largest pharmaceutical customers such as AstraZeneca and GSK and help provide solutions across, Operations, Supply Chain, Commercial, Central Functions and more recently R&Ds.

Insight and Visualization - Deliver Today, Not Tomorrow

In today's increasingly efficient and more transparent pharmaceutical company there is an increasing need to exploit all of the relevant data that sits both inside and outside of the firewall. Divisions and departments are no longer prepared to wait many months or years for huge projects that promise to deliver everything they might require but instead need and want to see results in a short space of time. See how QlikView can help answer your questions and deliver visual representations that people want to use.

James Mucklow, Member of the Management Group, PA Consulting



James helps clients deliver value from complex large scale programs. Over the last 16 years he has focused on the health sector and systems to improve patient care. He has experience from a range of sectors including health, finance, transport, energy, and government. This has been consolidated across all stages of the life – cycle from business concept to system implementation. James' key strengths are a deep understanding of technology, combined with a strong focus on delivery. Relevant experience includes delivery of the NIHR IS Systems Cloud Migration Program, NHS England's Integrated Intelligence Tool, Summary Care Record delivery of assurance and large scale programs. James sponsors the IT Architecture service within the IT Delivery practice. He is a Fellow of the British Computer

Society.

David Berry, Managing Consultant, Business Intelligence, PA Consulting



David is an expert in Business Intelligence (BI) and Information Management (IM) based on over a decade of experience in consulting, management, and business development. This experience is derived from enterprise customer engagements across Pharmaceutical, Financial Services, Insurance, Retail and Consumer Packaged Goods companies in Europe and the United States. David has developed numerous BI strategies and implementation roadmaps for complex businesses working with senior representatives. David has also project managed successful end to end implementations of complex global solutions, often working with teams based in multiple geographic locations and made up of multiple delivery partners.

Getting to Insights in RWE Faster Using Visualization and Cloud Technology

EMA GVP and FDA require ongoing signal detection against both external and internal safety databases. A failure to do this can have serious adverse consequences. It can also lead to missed re-use opportunities.

PA and QlikTech will demonstrate the first Clinical And Safety Signal technology (CLASS) to analyze clinical data sources from regulators, hospitals and industry to address this need. The technology evaluates clinical trials and safety data from over 283,000,000 patients reporting over 75,000 SAEs from over 144,000 studies from over 10,000 companies and research centers across the world.

Pek Lum, Chief Data Scientist, VP of Solutions, *Ayasdi*



Pek is the Chief Data Scientist and VP of Solutions at Ayasdi. Ayasdi is an enterprise software company that uses math, computer science, automation and domain knowledge to solve problems in precision medicine, drug discovery, biomarker identification, fraud-waste-abuse detection, hospital carepaths to name a few. She is responsible for the data science and solutions delivery team at Ayasdi where she provides leadership for the core research of the software product and data analytics strategies for customers. Prior to joining Ayasdi, Pek spent ten years applying genomics technologies and cutting-edge analytical methods at Rosetta/ Merck & Co. for the discovery of new targets and biomarkers for various therapeutic areas including cardiovascular diseases, diabetes and obesity. Pek obtained her PhD from University of Washington and her BS and MS from Hokkaido University in Japan.

Applications of Ayasdi Software to Precision Medicine, Drug Discovery and Biomarker Identification

The Ayasdi software platform is used in the life sciences field to solve problems in precision medicine, drug discovery and biomarker identification. The Ayasdi software platform combines a large repertoire of statistical and machine learning methods in a topological framework, allowing analysis of complex datasets large and small in a seamless and semi-automated manner. The foundation of the Ayasdi software platform is TDA (Topological Data Analysis) that has roots in topology, a branch of mathematics that dates back to the 18th century. Three characteristics of TDA that uniquely allow it to tackle complex data are the following: coordinate freeness, deformation invariance and compression. Coordinate freeness allows data collected from different platforms to be compared at a network level. Deformation invariance enables tolerance to noise. Lastly, compression allows complex data to be visualized as a simple network. The Ayasdi software platform has been used on a wide variety of data types such as next generation sequencing, imaging, microarrays, SNP, electronic medical records and clinical data. In this software demo, we will show how the Ayasdi software platform can be used to greatly facilitate various phases of drug discovery.

David Brown, CTO, Life Sciences, *BT Global Services*

David is responsible for developing and implementing the technical vision of BT Global Service's health & life science set of offerings. David joined BT in 2006 and brings over 25 years of experience in the health and life science industries. Prior to joining BT, David was a director of a company focused on providing disease management programs in the areas of cardiovascular disease, diabetes, and oncology. Prior to this, David held senior management positions in various companies including PwC, GSK and Digital Equipment Company.

A “Walled Garden” Collaborative Research Environment in the World of Translational Research

David Brown will present the work BT are carrying out to develop a cloud based walled garden for secure collaborative research between Pharmaceutical companies and healthcare providers. These environments will allow access to patient level phenotypic and genotypic information whilst control risks to patient confidentiality.

SESSION IV: KEYNOTE PRESENTATION

Chair: Matteo di Tommaso, VP, Research Business Technology, *Pfizer*

Professor Dame Janet Thornton, FRS, Director, *EMBL-EBI*



Since 2001 Janet has been Director of the European Molecular Biology Laboratory-European Bioinformatics Institute on the Wellcome Trust Genome Campus at Hinxton, near Cambridge, UK. EMBL-EBI provides core bioinformatics services, especially public biological data for genome and protein sequences and structures, small molecule data and integrated proteomic, metabolomics and transcriptome information. The institute also performs investigator led bioinformatics research.

Janet also has a research group specialising in protein structural bioinformatics and the computational biology of ageing. She has published more than 400 scientific papers, was elected to the Royal Society in 1999 and is a Foreign Associate of US National Academy of Sciences. In 2012 she became a Dame Commander of the Order of the British Empire for services to bioinformatics.

Janet has long interacted with colleagues in industry, firstly through the UK protein Engineering Club and case studentships and then as founder of Inpharmatica, an SME focused on the informatics of biology and chemistry, as applied to drug discovery. More recently she has strongly encouraged interaction of EBI staff with members of the EBI Industry Program and helped to establish the new Innovation and Translation initiative which seeks to build major pre-competitive collaborations with companies which utilize and benefit from EBI's research and service activities.

Janet coordinated the ELIXIR preparatory phase (2006–2013) and leads the commission-funded BioMedBridges project, which clusters the biomolecular sciences research infrastructures to provide data bridges and ensure data interoperability between them.

Visualizing the Value of Big Data at EBI

SESSION V: BRINGING IT ALL TOGETHER

Chair: Martin Leach, VP, R&D Information Technology, *Biogen Idec*

Round Table Discussions.

Table 1: P1 - Turquoise

Table 2: P2 - Orange

Table 3: P3 - Purple

Table 4: P4 - Red

Table 5: P5 - Yellow

Table 6: P6 - Green

Table 7: P7 - Brown

Table 8: P8 - Black

Table 9: P9 - White

SESSION VI

Chair: Dan Chapman, Senior Principal Scientist, Software Development and Architecture, *UCB*



Dan Chapman is part of the leadership team within Informatics at UCB with responsibility for Software Development and Architecture and Therapeutic Informatics (UK). Dan has 15 years experience working within the Pharmaceutical industry in a variety of roles.

After completing a PhD in Chemistry at Warwick University, Dan transitioned to informatics during post-doctoral research at Cambridge University as part of the CLIC consortium. Dan joined AstraZeneca in 1997 and worked on a variety of global projects before joining UCB in his present role in 2005. Since then, Dan has driven several projects to revolutionize the informatics platform within UCB and is currently actively involved in promoting Data Science across UCB.

David Belson, Senior Director, Industry & Data Intelligence, *Akamai Technologies*



David Belson is Akamai's Senior Director of Industry & Data Intelligence, responsible for Competitive Intelligence & Analysis, and data-driven evangelism, including Akamai's quarterly State of the Internet report series. A fifteen-year veteran of Akamai, Mr. Belson was one of Akamai's first Technical Consultants, and has also held a number of roles within Channel Sales, Product Marketing, and Product Management, where he was responsible for the launch of Akamai's Web Application Accelerator service. Prior to Akamai, Mr. Belson spent four years at GTE Internetworking/BBN, where he supported sales of one of the first enterprise-class managed Web hosting services, as well as working on an early prototype of what would currently be considered to be a SaaS platform. Mr. Belson has a Master's in

Technical and Professional Writing from Northeastern University in Boston, MA, and two Bachelor's Degrees (BS in Computer Science, BA in Science, Technology & Society) from Stevens Institute of Technology in Hoboken, NJ.

Visualizing the Internet: The Akamai Approach

Most people don't think much about the Internet beyond making sure their device is online, but with thousands of interconnected networks, hundreds of millions of users and billions of connected devices, the Internet is ultimately an extremely complex place. Akamai minimizes that complexity for customers through services providing performance, reliability, scalability, and security for customer Web sites, applications, and media. To be able to deliver these services, vast amounts of data are gathered as inputs into the Akamai service platform, with constant measurements across various metrics related to Internet performance, architecture, and interconnectedness. Customer usage of Akamai's services drive on the order of two trillion content requests per day, generating tremendous amounts of data, since every content request, both good and bad in nature, carries associated information that can provide unique insights at an aggregate level.

This talk will provide insight into the data collected and generated by Akamai, the visualizations that enable Akamai employees and customers to derive actionable value from the data, and case studies from biopharma customers.

NOTES
