PRISME Forum

Pharmaceutical R & D Information Systems Management Executives

PRISME Forum
TECHNICAL MEETING

Data-readiness in a World of AI

PRISME Forum Chair:

Olivier Gien, *Sanofi*

PRISME Forum Technical Meeting Chair:

Christian Baber, Head of R&D IT, *Shire*

November 14-15, 2018

Deerfield, IL

Host: Takeda

Download and stay up-to-date with our

PRISME Forum Fall 2018 Technical Meeting App:

http://my.yapp.us/PRISMETECH

Meeting Venue

The PRISME Forum Spring 2018 Technical Meeting will be hosted by Takeda and held at:
1 Takeda Pkwy, Deerfield, IL 60015, USA

Hotel

Hyatt Regency Deerfield, 1750 Lake Cook Rd, Deerfield, IL 60015.

Contacts

Program Coordinator:		+44.77.68.173.518/jcmwise@prismeforum.org
Secretariat	Office:	+1.224.938.9523/logistics@prismeforum.org
	On site, SMS:	+1.312.622.1234

PRISME Forum Technical Meeting Advisory Committee

Christian Baber (PRISME Forum Technical Meeting Chair), Head, R&D IT, *Shire*

Nick Brown, Head, AI and Data Science, *AstraZeneca*

Dan Chapman, Head, IT New Medicines Information Management, *UCB*

David Christie, Vice President, Enterprise Applications Group, *CSL Behring*

Lars Greiffenberg, Director R&D Information Research, AbbVie Library Sciences & Academic Partnerships, *AbbVie*

Carol Rohl, Executive Director, Global Research IT, *Merck*

Martin Romacker, Principal Scientist, Pharma Research Early Development Informatics, *Roche Innovation Center Basel*

Nico Stanculescu, Logistics Coordinator, *PRISME Forum*

Jianchao (JC) Yao, Head of Translational Medicine IT & SSF IT Lead, *Merck*

Jason Tetrault, Global Head Data Engineering and Emerging Technologies, *Takeda*

John CM Wise, Program Coordinator, *PRISME Forum*

PRISME Forum Host

The PRISME Forum Technical Meeting Advisory Committee would like to thank Takeda for hosting the 2018 PRISME Forum Fall 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."

Data-readiness in a World of AI

One of the key points of discussion at the last two PRISME Forum Technical Meetings on the topic of AI was that the limitations for AI/ML was not computing power, nor indeed algorithms, rather it was the availability of high-quality and fit-for-purpose structured data sets labeled both with appropriate metadata and endpoints. The scarcity of data for training machine learning is a fundamental feature of AI in the Life Science industry. Living systems are complex and noisy and as such require a significant amount of data to model them accurately. While substantial amounts of *in vitro* experimental data exist, *in vivo* data is much more difficult to collect and, in the case of human data, use is limited by informed consent, privacy regulations and ethical considerations.

The idea that 'data is more important than algorithms', has been gaining support since 2001 when Banko et al. published their paper "Scaling to Very Very Large Corpora for Natural Language Disambiguation"[i] which demonstrated that several very different Machine Learning Algorithms performed almost identically well on the complex problem of natural language disambiguation once they were given enough data.

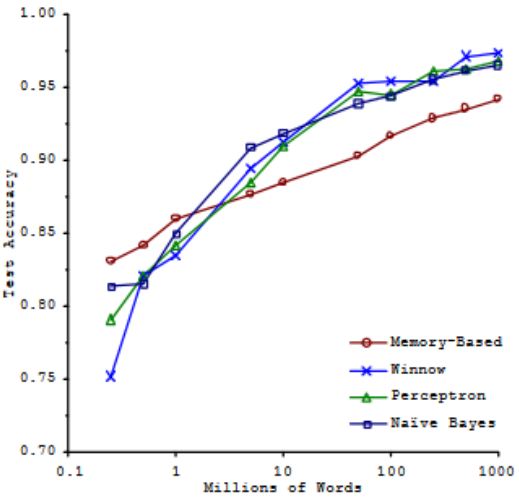


Figure 1. Learning Curves for Confusion Set Disambiguation

The idea was, more recently, taken up by an article entitled "The Unreasonable Effectiveness of Data"[ii] by Peter Norvig et. al. in 2009 which showed (Figure 1) that it can be relatively easy to reach around 50% accuracy using a variety of algorithms but to improve further, the need for data grows logarithmically. For AI to be effective a sufficient amount of high-quality data needs to be readily available.

The biopharmaceutical and healthcare industry in its entirety has a great deal of data. However, this data is rarely in a form amenable to use to train AI/ML methods without substantial data cleanup and labeling with meta-data and endpoints.

Additionally, this data is generally widely dispersed both within individual companies and between companies. This causes problems with gaining access to the data and, with the diversity of data formats, reading and understanding the data. Individual biopharmaceutical companies self-evidently have less data on which to train AI/ML systems to produce robust and generalizable results. If there were cross-company collaboration to merge data sets then much larger, more diverse and more effective training data sets could be made available. Despite this, the industry is cautious about sharing its data; not least because companies fear they will compromise or lose their IP. Other alternatives to address the issue include methods that mitigate data shortage and overfitting such as transfer learning, multi-task learning and the generation of synthetic data.

This PRISME Forum Technical Meeting will set out to explore opportunities for the biopharmaceutical industry to improve timely access to sufficient, high-quality data, on which AI systems can be trained (both within and beyond individual companies) and to use best the available data in the age of AI. A focus will be on practical examples that have been implemented at pharmaceutical companies along with efforts that have been attempted, but failed, and associated lessons learned.

Topics that will be addressed include:

- The implementation and use of the FAIR data principles (Findable, Accessible, Interoperable, Reusable)[iii] in industry
- Current tools and methods for meta data capture, end-state labeling and automated data preparation both at the point of creation and the time of use

- Practical storage, management and access to data from every stage of the R&D process and examples of data re-use & models constructed with data federated across multiple domains.
- Examples of the use of methods such as transfer learning to reduce the amount of directly relevant data required to build models for specific tasks.
- Methods that would allow companies to share their data, including the use of "guest-algorithms" that can train on data sets without exposing the IP.
- Identification of the most tractable domains within biopharma – both for internal development and where cross-industry data sets for AI training could be created.

[i] https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/acl2001.pdf
[ii] https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/35179.pdf
[iii] https://www.nature.com/articles/sdata201618

PROGRAM

PRISME Forum Technical Meeting sessions will be held at Takeda (1 Takeda Pkwy, Deerfield, IL 60015, USA).

WEDNESDAY, November 14, 2018			
19:00	*Welcome Reception at Hyatt Deerfield hotel (1750 Lake Cook Rd, Deerfield, IL 60015)*		
THURSDAY, November 15, 2018			
07:30	*Gather in the hotel lobby for departure to the meeting venue*		
08:00	*Check-in, continental breakfast, and poster installation*		
08:30	**Welcome Notes & Introductions**		**Olivier Gien,** VP, Global Head Medical IT, *Sanofi*; Chair, *PRISME Forum* **Christian Baber,** Head, R&D IT, *Shire;* Technical Meeting Chair, *PRISME Forum*
SESSION I A: PLENARY PRESENTATIONS			**Chair: Christian Baber, Head, R&D IT, *Shire***
08:50	**Plenary #1 – Data for Drug Discovery**		**John Overington,** CIO, *Medicines Discovery Catapult*
09:30	**Plenary #2 – Data First - Information Procurement based on FAIR Data**		**Martin Romacker,** Principal Scientist, *Pharma Research Early Development Informatics, Roche Innovation Center Basel*
10:00	*Coffee Break*		
SESSION II: START-UP COMPANY "PITCH" SESSION			**Chair: David Christie, Vice President, Enterprise Applications Group, *CSL Behring***
10:25	*Introduction*		
10:30	**1**	**Startup #1 – Outlier** *Automated and Useful Insights from Your Data, Personalized to Your Needs*	**Sean Byrnes**, CEO, *Outlier*
10:45	**2**	**Startup #2 – HealthVerity** *Activating Clinical Data for AI*	**Andrew Goldberg,** COO, *HealthVerity*
11:00	**3**	**Startup #3 - Healthdata.link** *Increasing the Availability of High-Quality Data*	**Jake Plummer,** CEO, *Health Data Link*
11:15	**4**	**Startup #4 – PathAI** *Artificial Intelligence for Pathology: From Discovery to Companion Diagnostics*	**Andy Beck,** CEO and Co-founder, *PathAI*
11:30	**5**	**Startup #5 – rMark Bio** *Data Preparedness for Machine Learning*	**Jason Smith,** CEO & Co-founder**,** *rMark Bio*
11:45	***Conclusions***		
	PRISME Forum Start-up Company "Pitch" Session Evaluation Panel Members: **Carol Rohl,** Global Research IT Head, *Merck* **Nick Brown,** Head of AI and Data Science**,** *AstraZeneca* **Martin Romacker,** Principal Scientist, *Pharma Research Early Development Informatics, Roche Innovation Center Basel* **Jason Tetrault,** Global Head, Data Engineering and Emerging Technologies, *Takeda* **Jianchao (JC) Yao,** Head of Translational Medicine IT & SSF IT Lead, *Merck*		

THURSDAY, November 15, 2018 (cont.)			
SESSION III A: POSTERS			**Chair: Dan Chapman, Head, IT New Medicines Information Management, *UCB***
11:55	*Introduction*		
12:00	*Poster Rotations (Three 15-minute rotations)*		
	P1	**DISQOVER**	**Hans Constandt**, CEO, *ONTOFORCE* **Filip Pattyn**, Scientific Lead, *ONTOFORCE*
	P2	**Quantum Molecular Design (QMD)**	**Ed Addison,** CEO, *Cloud Pharmaceuticals*
	P3	**Lab Data to Machine Learning in 30 seconds in PK/Tox and Solid Dose Formulations**	**Timothy Rhodes,** Senior Investigator, *Merck* **Tim Gardner,** CEO & Founder, *Riffyn*
	P4	**Partnering around ADMET Data**	**Guido Lanza,** CEO, *Numerate*
	P5	**Data-readiness in a World of AI**	**Narayanan Ramaswamy,** Global Director Solutions Architecture & Technology Enablement**,** *Otsuka*
	P6	**Measuring and Optimizing Time to Data**	**Tim Delisle,** CEO, *Datalogue*
12:45	*Lunch*		
SESSION III B: POSTERS			**Chair: Dan Chapman, Head, IT New Medicines Information Management, *UCB***
14:00	*Poster Session (Remaining three 15-minute rotations)*		
SESSION I B: PLENARY PRESENTATIONS			**Chair: Christian Baber, Head, R&D IT, *Shire***
14:45	**Identifying and Exploiting Diverse and Difficult Datasets for Training ML**		**Andrew Carroll,** Product Lead, *GoogleBrain, GoogleAI*
15:15	**CEDAR: Semantic Technology in Support of Open Science and Improved Knowledge Management**		**Mark Musen,** Professor, Biomedical Informatics and Biomedical Data Science, *Stanford University*
SESSION IV: BREAK-OUTS, READOUTS & COFFEE			**Chair: Lars Greiffenberg, Director R&D Information Research, AbbVie Library Sciences & Academic Partnerships, *Abbvie***
15:45	**Breakout Session –** Members and meeting guests will be divided into four groups led by co-captains		**Co-captains:** **Group A (Start-up Pitches): Carol Rohl** and **Sean Byrnes** **Group B (Plenaries 1&2): Nick Brown** and **Andrew Goldberg** **Group C (Posters 1-3): Martin Romacker** and **Jake Plummer** **Group D (Posters 4-6): Jason Tetrault** and **Andy Beck** **Group E (Plenaries 3-4): Jianchao (JC) Yao** and **Jason Smith**
16:15	**Plenary Session –** Readouts from breakout groups		
SESSION V: MEETING SUMMARY, AWARDS & RECEPTION			**Chair: Christian Baber, Head of R&D IT, *Shire***
16:45	Meeting Summary		
17:00	Awards & Networking Reception		
18:00	*Return to the hotel*		
19:00	*Informal dinner (gather in the hotel lobby for transfer to restaurant)*		

BIOS AND ABSTRACTS

PRISME Forum Chair: Olivier Gien

VP, Global Head Medical IT, *Sanofi*

Olivier Gien, PhD, is a Chemical Engineer by training and holds a PhD in Organic Chemistry. His PhD work focused on leveraging Artificial Intelligence technologies and retro-synthetic analysis to build a system helping chemists in the design of synthetic routes.

Dr. Gien started his career in the Exploratory Unit of Sanofi's Hungarian affiliate in Budapest then took charge of Information Systems for Industrial Chemical development at Sanofi's Sisteron site. He then led Global Discovery Research Information Systems at Sanofi-Synthelabo, followed by Sanofi-Aventis in Montpellier, before taking on the roles of Global Head, R&D IT in 2010, Global Head, Clinical IT in 2015 and finally, Global Head, Medical IT in 2017.

PRISME Forum Technical Meeting Chair: Christian Baber

Head, R&D IT, *Shire*

Christian Baber, PhD, is a chemist by training and holds undergraduate and PhD degrees in computational chemistry with a focus on AI techniques to assess the synthetic accessibility of de novo design compounds. Christian continued this work with a post-doctoral fellowship on the automated design of targeted combinatorial libraries at the Department of Knowledge Engineering, Osaka University, Japan before moving into industry as a computational chemist and cheminformatician.

Christian has a wide breadth of experience across companies ranging from startups to Pfizer and diverse therapeutic areas with a focus on early stage lead identification and screening. Christian has been with Shire since 2015 and is currently the Head of R&D IT where he leads IT efforts for the worldwide Research and Development organizations, and is in the process of building out R&D Informatics after doing the same for the discovery functions. Prior to Shire, Christian was the Head of Cheminformatics and Compound Management and Data Steward at Cubist Pharmaceuticals where, amongst other things, his team was responsible for automation, high-throughput screening, scientific programing and the corporate scientific database.

INTRODUCTORY NOTES

SESSION IA: PLENARY PRESENTATIONS

Session Chair: Christian Baber

Head, R&D IT, *Shire*
Technical Meeting Chair, *PRISME Forum*

John Overington

CIO, *Medicines Discovery Catapult*

John Overington, PhD, studied Chemistry at the University of Bath, followed by a PhD at Birkbeck College. There he developed automated approaches to protein modelling, and explored protein sequence-structure relationships. He then held a postdoctoral position at the Imperial Cancer Research Fund (now part of CRUK). John then joined Pfizer (1992), originally as a computational chemist, progressing to a role where he led a large and highly multidisciplinary group combining rational drug design with structural biology.

In 2000 John moved to the biotech company, Inpharmatica, where he led the development of a series of computational and data platforms to improve drug discovery; these included the medicinal chemistry database StARLite. In 2008 John was central to the transfer of this database to the EMBL-EBI, where the successor is now known as ChEMBL. While there his work expanded into large-scale patent informatics with the Open patent database SureChEMBL. John then moved (2015) to a London-based Artificial Intelligence technology company - Stratified Medical (later renamed BenevolentAI), where he continued his translational drug discovery informatics R&D activities, applying machine learning methods to the development of novel biomedical data extraction and integration strategies.

In 2017 John joined the UK's Medicine Discovery Catapult as CIO, where he leads the development and application of informatics approaches to promote and support innovative, fast-to-patient drug discovery in the UK through collaborative projects across the applied R&D community.

Data for Drug Discovery

John Overington

Martin Romacker

Principal Scientist, *Pharma Research Early Development Informatics, Roche Innovation Center Basel*

Martin Romacker, PhD, is a Data and Information Architect in Technical Solution Delivery and Architecture in Pharma Research and Early Development Informatics at F. Hoffmann-La Roche. His main areas of activities are terminology management, semantic engineering, scientific data curation, text mining and semantic search.

Martin has been active in the development and deployment of Data Standards within the organization but also teaming up with external partners. During his work Martin puts a particular emphasis on cross-functional data organization and the definition of Data Quality KPIs. As a part of pREDi's precompetitive engagements Martin has been contributing to projects of the Pistoia Alliance and the Innovative Medicine Initiative.

Before joining Roche, Martin worked at the Novartis Institute of Biomedical Research as a Senior Knowledge Engineer Consultant leading various ontology development and text mining projects such as the implementation of BioAssay Ontologies. Martin was also Managing Director of a Start-Up company in Freiburg (Germany) where he focused on R&D activities for NLP.

He holds a PhD in Computational Linguistics from the University of Freiburg on Semantic Interpretation of technical and medical texts.

Data First - Information Procurement Based on FAIR Data

Martin Romacker

The last 6-18 months brought a significant change to the pharma industry. There is a growing consensus amongst executives that data need to be considered and, therefore, be treated as a – if not the most – precious asset of the company. A couple of reasons hold accountable for this tremendous change of perception. New players have entered the scene not taking a lab-based but a data-based approach combining data access at scale with AI-based algorithms and high-performance computing. Translational approaches, Precision Medicine and Personalized Health-Care require a data management strategy breaking up traditional data silos within the organization spanning a broad range from research, pre-clinical, clinical, diagnostic and real-world data.

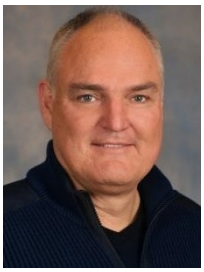
In contrast to the urgent need to bring in high-quality data into advanced analytics and data science, so far the pharma industry has done a poor job on a cross-functional data management strategy. As a consequence, significant resources have to be allocated for data acquisition, data normalization and data integration although absolutely no value is created by this kind of activities. Organizational, technological and semantic thresholds are still high to enable fast and agile data sharing. Finally, the application of AI suffers from poor data quality and outcomes are not always reliable. Remediation happens at local level so that data curation and cleansing exercises are repeatedly performed over time, across the company and even in the industry.

The process of acquiring and processing information should be defined as any other procurement process with clear specification of the data types, metadata and terminologies/ontologies. Information parts and components should seamlessly fit together and integration overhead needs to be reduced by at least 50 % in terms resources and 90 % in terms of time to boost the productivity of our industry. The FAIR principles (findable, accessible, interoperable and reusable) give a very clear guidance how this could be achieved. For the IT part of the journey we need to build technical capabilities, collaborate across pre-competitive initiatives to establish standards and last but not least promote data governance.

SESSION II: Start-Up Company "Pitch" Session

CHAIR: David Christie

Vice President, Enterprise Applications Group, *CSL Behring*



At CSL Behring, **David Christie** works with business function leaders and the CIO to provide operational and strategic leadership in support of CSL Behring's global business strategies.

Previously, David was Vice President, Research and Development Informatics at Amgen. Prior to that, he led Amgen's Global Commercial Operations and Corporate Functions IS group. David has also led Amgen's International IS function based out of Zug, Switzerland, and before that, the IS group supporting Global Development functions.

Earlier in his career, David worked at Eli Lilly in various roles in Australia, New Zealand, and the US. David holds a Bachelor of Business from the University of Technology, Sydney.

The session's objective is to provide a constructive yet relaxed activity to encourage interaction between PRISME Forum members and five start-up companies with a value proposition relevant to the context of the meeting's theme, i.e., Data-readiness in a world of AI. The rationale, in particular, is that:

- *the PRISME Forum members get introduced earlier than they otherwise would to relevant life science R&D/healthcare AI-based business propositions.*
- *the start-up companies have an opportunity to interact with members of the PRISME Forum and get some constructive feedback about their business propositions.*

In terms of structure, the session will begin with the co-chairs' overview followed by five 15-minunte "pitches" delivered by the five start-ups showcased below. Each of these four segments will allow the Panel to ask questions. The session will end with the co-chairs' summary and concluding notes.

Start-up Company "Pitch" Session Evaluation Panel Members

Carol Rohl, Global Research IT Head, *Merck*

Carol Rohl, PhD, is responsible for information technology and data strategy and solutions for Discovery, Preclinical and Early Discovery in MRL. She joined Merck and Co in 2005 as a member of the Rosetta Inpharmatics informatics group where she was initially responsible for pathway centric informatics analysis capabilities. Subsequently she became director of the molecular informatics group within the molecular profiling and research informatics department in Merck Research labs, with responsibility for genomic information systems and informatics capabilities in support of profiling, genomics and biomarker efforts. In 2010, she moved to MRL IT as part of the Informatics organization where she was responsible for translational informatics and built a health and clinical informatics capability within Informatics.

Prior to her current role, she established and led the Scientific Information Management within MRL IT, focused on advancing our maturity around effective management and use of information. Prior to joining Merck, Rohl was an assistant professor of biomolecular engineering at the University of California, Santa Cruz where she led a research team focused on the development and application of the Rosetta protein structure prediction algorithm to problems in protein design and protein fold evolution.

Dr. Rohl earned a PhD in biochemistry from Stanford University in the laboratory of Dr. Robert Baldwin, where her thesis work focused on the peptide model systems for protein folding and stability. Additionally, she did postgraduate work in the laboratories of Dr. Rachel Klevit and Dr. David Baker and the University of Washington in Seattle.

Nick Brown

Head of AI and Data Science, *AstraZeneca*

Nick Brown is the Head of Data Science & AI representing the science units with AstraZeneca IT. He leads three teams that apply data science approaches in Early Science, Late Science and Knowledge Management areas. His main areas of interest are knowledge engineering, scientific big data analytics, semantic search, knowledge graphs and AI web services.

He holds a Genetics BSc and Bioinformatics Masters at York University where he created neural networks and hidden markov models for predicting transmembrane proteins. He initially worked for the Forensic Science Service identifying people from the DNA of touched objects before joining AstraZeneca in 2001 originally as a computational toxicologist.

Nick has held leadership roles in R&D and IT for the past 15 years where he led development & deployment of a fully automated computational image analytics platform for high throughput screening, a next generation search platform for the company and creation of a semantic analytics engine for drug repositioning. In the last few years, Nick led the Technology Innovation Lab for the CTO working across the organization to scout and evaluate new, emerging technologies to critical business problems in R&D, Operations and Commercial. He's been lucky enough to work with start-ups using advances in AI, ML, IoT, blockchain, edge compute, AR, VR and several other IT acronyms!

Martin Romacker

Principal Scientist, *Pharma Research Early Development Informatics, Roche Innovation Center Basel*

Martin Romacker, PhD, is a Data and Information Architect in Technical Solution Delivery and Architecture in Pharma Research and Early Development Informatics at F. Hoffmann-La Roche. His main areas of activities are terminology management, semantic engineering, scientific data curation, text mining and semantic search.

Martin has been active in the development and deployment of Data Standards within the organization but also teaming up with external partners. During his work Martin puts a particular emphasis on cross-functional data organization and the definition of Data Quality KPIs. As a part of pREDi's precompetitive engagements Martin has been contributing to projects of the Pistoia Alliance and the Innovative Medicine Initiative.

Before joining Roche, Martin worked at the Novartis Institute of Biomedical Research as a Senior Knowledge Engineer Consultant leading various ontology development and text mining projects such as the implementation pf BioAssay Ontologies. Martin was also Managing Director of a Start-Up company in Freiburg (Germany) where he focused on R&D activities for NLP. He holds a PhD in Computational Linguistics from the University of Freiburg on Semantic Interpretation of technical and medical texts.

Jason Tetrault

Global Head, Data Engineering and Emerging Technologies, *Takeda*

Jason Tetrault is the Global Head Data Engineering and Emerging Technologies at Takeda. He is technologist with 18 years of experience in Software and Systems development from large scale systems to mobile devices.

He has spent the past 7 years in the biopharmaceutical industry with a focus on R&D, Genetics and Genomics, Big Data and Data Science. Before Takeda Jason has led teams and initiatives around Data Science, Big Data and Cloud Scaling at Biogen and AstraZeneca.

Jianchao (JC) Yao

Head of Translational Medicine IT & SSF IT Lead, *Merck*

JC Yao, PhD, MS, is the Director, Head of Translational Medicine and Oncology IT and IT lead for Merck's South San Francisco Discovery Site.

As the IT Client Service Leader, he is accountable for technology and information sciences capabilities to enable Translational Medicine and Oncology business objectives.

As the South San Francisco Discovery Site IT lead, he is responsible to understand site-based issues and opportunities relevant to IT, informatics and data science and accountable for technology and information sciences capabilities to realize the site mission and vision.

JC joined Merck Research Laboratory IT in 2013 after completing a postdoctoral fellowship at Cold Spring Harbor Laboratory. He earned his PhD in Molecular Biology with a focus on Bioinformatics from the University of Texas at Austin in 2009, where he also received a Master's degree in Statistics and a Certificate in Business from the McCombs School of Business.

Start-up #1: Outlier

Sean Byrnes

CEO, *Outlier*

Sean Byrnes, MEng, is the CEO of Outlier.ai, a new company that is reinventing business intelligence using artificial intelligence.

Prior to Outlier, Sean was the founder of Flurry, the leader in advertising and analytics services for mobile applications acquired by Yahoo! in 2014.

He is also an advisor, mentor and angel investor in the San Francisco bay area. Sean earned an undergraduate engineering degree from Dartmouth College and a Master of Engineering in Computer Science from Cornell University.

Automated and Useful Insights from Your Data, Personalized to Your Needs

Sean Byrnes

Outlier is a new way of thinking about data analysis. Instead of requiring intensive manual exploration of data to find value, Outlier uses unsupervised machine learning techniques to automatically find the valuable patterns hiding in your data. After spending only a few minutes connecting Outlier to your data, you will immediately begin to receive useful and insightful insights from your data, personalized to your needs. Outlier is already working with some of the largest businesses and largest data sets to find insights every day, spanning the pharmaceutical, travel, media and e-commerce industries.

Start-up #2: HealthVerity

Andrew Goldberg

Chief Operating Officer, *HealthVerity*

Andrew Goldberg is the co-founder and COO of HealthVerity where he is responsible for the day-to-day operations and performance excellence of the company. He was formerly the SVP Strategy and Marketing for Dialogic, responsible for corporate development and global marketing, which he sold to Novacap.

He was previously the Vice President, Corporate Development for Avaya which was sold to Silverlake/TPG. Andrew also co-founded Eziaz, a venture-backed broadband access company, and held various roles at Comcast, Bain & Company and Diageo.

Andrew holds an MBA with Distinction from Harvard Business School and a BA with Honors from the University of Pennsylvania.

HealthVerity: Activating Clinical Data for AI

Andrew Goldberg

The HealthVerity technology platform serves as the foundation for the rapid creation, exchange and management of healthcare and consumer data in a fully-interoperable, privacy-protecting manner. Advantaged by highly sophisticated identity resolution and data linking capabilities, HealthVerity is on a mission to increase transparency and activate deeper insights across the healthcare industry.

Start-up #3: Health Data Link

Jake Plummer

Chief Executive Officer, *Health Data Link*

Jacob (Jake) Plummer is CEO for Health Data Link, a company creating the first global identity solution in healthcare. Previously, he was co-head of US new business and head of global business development for Allscripts, the third largest health IT company in the world.

Jake actively supports education programs in STEM initiatives and serves as President of the Illinois Mathematics and Science Academy Foundation.

He completed graduate work in business and public policy at The University of Chicago and majored in economics at Knox College.

Increasing the Availability of High-Quality Data

Jake Plummer

Health Data Link enables an ecosystem where a variety of patient data can be assembled into de-identified, de-duplicated, and enriched data sets to drive research and patient care. Developed in a clinical setting with nearly a decade of real-world patient data, HDL's highly-secure solution enables collaboration between stakeholders with sensitive data including academic medical centers, payors, government organizations, and public health institutes.

Start-up #4: PathAI

Andy Beck

Chief Executive Officer and Co-founder, *PathAI*

Andy Beck, MD, earned his MD from Brown Medical School and completed residency and fellowship training in Anatomic Pathology and Molecular Genetic Pathology from Stanford University.

He completed a PhD in Biomedical Informatics from Stanford University, where he developed one of the first machine-learning based systems for cancer pathology. He's been certified by the American Board of Pathology in Anatomic Pathology and Molecular Genetic Pathology. Prior to co-founding PathAI, he was on the faculty of Harvard Medical School in the Department of Pathology at Beth Israel Deaconess Medical Center. He has published over 110 papers in the fields of cancer biology, cancer pathology, and biomedical informatics.

Artificial Intelligence for Pathology: From Discovery to Companion Diagnostics

Andy Beck

Pathologic analysis of patient tissue specimens plays a central role in the field of oncology. Recent advances in artificial intelligence and computer vision offer tremendous potential for discovering new pathologic mechanisms of cancer treatment response and identifying new diagnostics for matching patients and therapies. We will discuss these new advances and their potential for accelerating progress in the diagnosis and treatment of cancer.

Start-up #5: rMark Bio

Jason Smith

Chief Executive Officer and Co-founder, *rMark Bio*

Jason Smith is the co-founder and CEO or rMark Bio. rMark Bio's patented intelligence platform, Fabric, unites the real-time business objectives of a pharmaceutical company with the current activities of academic and clinical researchers to deliver evidence-based decisions that remove inefficiencies and increase the ROI in thought leader engagements.

Prior to rMark Bio, Jason held positions in early-stage companies (xSides, Cryptocybernetics), large multinational corporations (IBM, ATI Research) and venture capital incubators (BE Labs) yielding a rich professional background. In his roles such as VP of Corporate Development, VP of Product, and Chief Architect he has been responsible for the management of various functional areas including go-to-market strategy, product development and corporate operations.

Data Preparedness for Machine Learning

Jason Smith

Fueled by data, machine learning and artificial intelligence will continue to drive innovation, growth and value within the Pharmaceutical Industry. However, what happens when the data is siloed, incomplete, biased, or otherwise unprepared for machine learning models to provide the value you need? Jason Smith, CEO of rMark Bio, will discuss of how they have helped Pharmaceutical IT, Analytics and Data teams overcome obstacles of access, cleanliness and formatting in preparation of machine learning analysis.

SESSION III: POSTERS

Session Chair: Dan Chapman

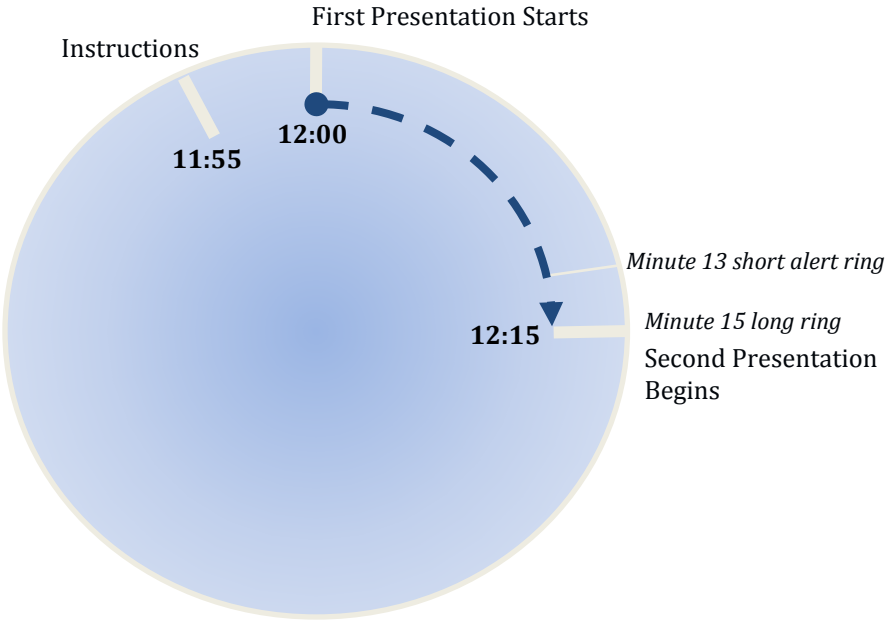
Head, IT New Medicines Information Management, *UCB*

Dan Chapman, PhD, leads the IT New Medicines Information (NMIM) Management group at UCB.

Dr. Chapman's team is driven to create an enabling environment of data that will help the scientific community derive better insights and promote data driven decisions. A core belief of the NMIM team is that the use of innovative technologies will lead to improved processes and support more efficient research. The responsibilities of the NMIM team include data capture platforms (ELN, Registration systems, LIMS, Logistics, Assay Data, MDM), Data Architecture (Responsible for integrating both out internal and external data to ensure that the right data is available to UCB scientists) and a team of data scientists providing bespoke tools and visualizations aimed at providing insights to UCB scientists.

Dr. Chapman has led the IT aspects of a number of acquisitions including emerging technologies new to UCB. He has over 20 years' experience working within the Pharmaceutical industry in a variety of Informatics and IT roles. He received his PhD in Chemistry from Warwick University and transitioned to informatics during post-doctoral research at Cambridge University as part of the CLIC consortium.

POSTER SESSION ROTATIONS



INSTRUCTIONS @11:55am

Session Chair will invite participants to take their seats.

Chair will open the session by introducing the presenters, their posters, along with the session structure and flow.

PRISME Forum staff will ring the bell on each rotation (minute 13 of each presentation and then minute 15 at which time the presentation must end).

Delegates are invited to identify the color of their lanyard and match to rotation called out by staff: "Rotation 1, 2, …n".

Rotations will involve shift of participants' groups to the next poster on their right.

FIRST SET OF ROTATIONS:

ROTATION 1 – 12:00	ROTATION 2 – 12:15	ROTATION 3 – 12:30
P1 – Orange	P1 – Green	P1 – Yellow
P2 – Gray	P2 – Orange	P2 – Green
P3 – Red	P3 – Gray	P3 – Orange
P4 – Purple	P4 – Red	P4 – Gray
P5 – Yellow	P5 – Purple	P5 – Red
P6 – Green	P6 – Yellow	P6 – Purple

BREAK FOR LUNCH

SECOND SET OF ROTATIONS:

ROTATION 4 – 14:00	ROTATION 5 – 14:15	ROTATION 6 – 14:30
P1 – Purple	P1 – Red	P1 – Grey
P2 – Yellow	P2 – Purple	P2 – Red
P3 – Green	P3 – Yellow	P3 – Purple
P4 – Orange	P4 – Green	P4 – Yellow
P5 – Gray	P5 – Orange	P5 – Green
P6 – Red	P6 – Grey	P6 – Orange

POSTER ROTATIONS *(lanyard colors)*

ORANGE	
Andy Beck	*Path AI*
Alastair Binnie	*BMS*
Nick Brown	*AstraZeneca*
Kelly Caruso	*Shire*
Jake Plummer	*Health Data Link*
Jean-Luc Schmidt	*Sanofi*
Michael Shanler	*Gartner*
JC Yao	*Merck*
RED	
Christian Baber	*Shire*
Edsel Calliste-David	*Astellas*
Andrew Carroll	*GoogleAI*
David Christie	*CSL Behring*
Andrew Goldberg	*HealthVerity*
Brian Martin	*AbbVie*
John Overington	*Medicines Discovery Catapult*
Carol Rohl	*Merck*
Jon Stevens	*AbbVie*
PURPLE	
Sean Byrnes	*Outlier AI, Inc.*
Thomas Frei	*Novartis*
Klaus Hofenbitzer	*Celgene*
Martin Leach	*Alexion*
Mike Montello	*GSK*
Arun Nayar	*Amgen*
Martin Romacker	*Roche*
Errol Sandler	*PRISME Forum*
Tatsuyuki Takahashi	*Mitsubishi Tanabe*
Ashok Upadhyay	*Otsuka*

YELLOW	
Pete Dhillon	*Daiichi-Sankyo*
Joel Ekstrom	*Ionis*
Martin Erkens	*Roche*
M. Hall Gregg	*Pfizer*
Phil Hayduk	*AbbVie*
Jay Krishna	*Shire*
Francois Midili	*Ferring*
David Sedlock	*Takeda*
Jason Smith	*rMark Bio, Inc.*
Deep Vaswani	*Astellas*
GREEN	
Andrew Allen	*Regeneron*
John Apathy	*Celgene*
Michael Cassidy	*Regeneron*
John Conway	*AstraZeneca*
Massimo de Francesco	*UCB*
Lars Greiffenberg	*AbbVie*
Hongmei Huang	*Genentech*
Scott Oloff	*Boehringer Ingelheim*
Susie Stephens	*Pfizer*
Jason Tetrault	*Takeda*
GREY	
Dan Chapman	*UCB*
Roy Ladd	*AbbVie*
Bruno Larmurier	*Servier*
Tomoyuki Matsunaga	*Takeda*
Natalie Mirutenko	*Takeda*
Mark Musen	*Stanford University*
Andy Newsom	*CSL Behring*
Leonard Sagalov	*AbbVie*
Etzard Stolte	*Roche*

Hans Constandt

CEO, *ONTOFORCE*

P1

Hans Constandt, MS, has a bachelor in medicine, a master in biotechnology, a software engineer degree and a master in innovation & entrepreneurship. He has +15 years experience in bioinformatics, software engineering and data architectures where he worked 12 years in a pharma multinational, Eli Lilly, as account manager, data architect, senior business consultant and global lead in knowledge management, data science and integrative informatics.

Hans (co-)founded 3 companies and is active as an advisor in other startups and scaleups. He is an avid speaker, passionate about what he does and he received many recognitions and awards for bringing disruptive technologies to market and successfully raising funds scaling up his company, ONTOFORCE.

His goal is to unleash the power of linked data on very large scales empowering citizen data science by democratizing access to data, for everyone. His vision is that linked data at work will impact everyone's life, starting in life sciences and healthcare healing patients with smarter data but also significantly reducing (research) cycles in small and big companies enabling them to fail early and bring products to market much faster.

P1: DISQOVER: A Paradigm Shift in Semantic Data Cataloging and Enabling AI with a Scalable, Transparent and Easy to Use New Semantic Data Ingestion Engine

Filip Pattyn, Hans Constandt

Semantic web technologies are gaining renewed interest since the technique of data indexing, layered on top of a traditional semantic triplestore, has been developed. This greatly improved the speed of the semantic applications and opened new opportunities.

DISQOVER is a web-based semantic search and exploration platform for linked data sources. The platform allows to ingest and harmonize a wide spectrum of public, private and third-party data which are glued together via an overarching DISQOVER configuration ontology. The system supports data federation between different DISQOVER installations and is capable to prepare and create visual analytics dashboards directly on top of the data. We will zoom in on binning and quantifying semantified data resulting in lightning fast visual analytics. Slicing and dicing a multitude of data sets in an easy way became simple and last month we released a plugin framework where (downstream and integrated) AI is used for several purposes including optimizing data curation, NLP, data ingestion, image analysis and protein structure matching. This poster presentation will introduce these newest developments focusing on lowering the threshold for semantically integrating, enriching and linking new data sources.

ONTOFORCE filed a patent on this new data ingestion engine as this is a novel concept in the space and allows data conversion processes to not only be managed in a visually attractive web-based application but eliminates the need to write and maintain conversion scripts. This new semantic data ingestion framework comes with a huge gain in speed, stability and scalability and includes a componentized semantic ETL engine where everyone can easily review, adjust and monitor semantically-enabled data ingestion pipes in a transparent way, thereby addressing the three major challenges in most of the traditional triplestores.

This plugin and pipeline architecture strategically fits the need to enhance the ability of bio-pharma to harness its data assets better and more easily to feed their AI programs. A typical biopharma use case will be presented integrating a set of data and metadata to create a data catalog of research data and ingestion into an AI component for structure search.

Ed Addison

CEO, *Cloud Pharmaceuticals*

P2

Ed Addison, MS, MBA, co-founded Cloud Pharmaceuticals in 2011. Addison is a serial entrepreneur who founded three previous ventures, two of which successfully merged with public companies. He has a unique and strong blend of in-depth business and technical experience in biotechnology and in information technology.

Ed holds an MS in Biomedical Engineering from Johns Hopkins University and MBA from Duke University Fuqua School of Business, including a diploma in Duke's Health Sector Management Program.

From 2000-2009, he focused on a series of ventures in the pharmaceutical and life sciences space. These included start-ups in the life sciences market, including BioFortis, a LIMS venture in Baltimore, and Inclinix, a CRO in Wilmington, NC, and TeraDisc, the predecessor of Cloud Pharmaceuticals, where he was named "Coastal Entrepreneur of the Year" in 2009. Ed has been working in drug discovery and development for over 12 years.

As a life sciences business development professional, he has negotiated numerous licenses for therapeutics and technology. He spent the early part of his career in information technology and has been an adjunct professor for over 25 years. Previously he taught at Johns Hopkins University where he developed courses in bioinformatics and entrepreneurship. He was the co-founder and program director of the online MS degree in Bioinformatics for Johns Hopkins University's Whiting School.

Ed received prior awards for entrepreneurship including "Entrepreneur of the Year" in 1994 by the Information Industry Association, and #51 on the National Fast 500, and Coastal Entrepreneur of the Year in 2009. He is the author of a book called Leveraging the Horizon about seed-stage technology ventures.

P2: AI Platform for Drug Design

Ed Addison

Cloud Pharmaceuticals has developed an AI and In Silico platform for the design of drug lead compounds that integrates multiple methods. Our platform, called Quantum Molecular Design (QMD), combines quantum chemistry, cloud computing and AI. QMD finds completely novel molecules from a search of "hot spots" illuminated in chemical space that have predictably high binding affinity, low toxic side effects and acceptable molecular properties using a "multi objective optimization' algorithm. AI is used to augment best of breed computational chemistry, by applying selected machine learning algorithms to appropriate data sets.

This presentation will focus on several AI methods used by QMD for predicting chemical properties, and the data sources that are used. These methods have been heavily tested in partnership the University of Florida, and are now being applied and testing in collaboration with GSK's new in silico discovery group. In particular, the following table illustrates some of the data sources to be discussed:

AI ALGORITHM DATA SOURCES USE

Heuristic Search of Molecular Space Data from prior attempts to adjust parameters for convergence
Prediction of toxic side effects Tox21 and private data sources
Prediction of binding affinity BindingDB and private binding affinity data sources
Prediction of Pk (in development) Atom (pending) and private data sources
Blood Brain Barrier Permeability Data extracted from literature

In addition to the multi property assessment methods for lead design above, Cloud is now beginning to apply machine learning to discover novel targets. A partnership has been formed with Parallel Profile, a new PGx clinical testing and assessment company, where Parallel Profile provides de-identified pharmacogenomics data to Cloud and machine learning methods are being used to identify novel targets. This project will be briefly summarized, illustrating how PGx data is being applied to target discovery.

Tim Gardner

CEO and Founder, *Riffyn*

P3

Tim Gardner, PhD, is the Founder and the CEO of Riffyn. He was previously Vice President of Research & Development at Amyris, where he led the engineering of yeast strain and processes technology for large-scale bio-manufacturing of renewable chemicals. Earlier, he was an Assistant Professor of Biomedical Engineering at Boston University, the Founder of Cellicon Biotechnologies, and a Programmer at ALK Associates.

Tim has been recognized for his pioneering work in Synthetic Biology by Scientific American, the New Scientist, Nature, Technology Review, and the New York Times. He also served as an advisor to the European Union Scientific Committees and the Boston University Engineering Alumni Advisory Board.

He holds BS in Mechanical Engineering from Princeton University and a PhD in Biomedical Engineering from Boston University.

P3: Lab Data to Machine Learning in 30 seconds in PK/Tox and Solid Dose Formulations

Tim Gardner

Today's research and development breakthroughs lie deep in the midst of complex, multivariate data sets. Yet, experimental anomalies and fundamental discoveries often go unnoticed due to data fragmentation and poor data quality. Insights are buried in uninterpretable spreadsheets, inaccessible databases, or excessive experimental noise.

The cloud-based Riffyn SDE software automatically links experimental designs to measurement data, and structures it for machine learning within seconds after it is collected in the lab. We will illustrate how Merck is using this capability to identify unexpected anomalies in animal PK/Toxicology studies across multi-site, multi-investigator environments, and to identify multivariate relationships between process parameters and product quality in oral solid dose formulations.

Guido Lanza

CEO, *Numerate*

P4

Guido Lanza is the CEO of Numerate, which he co-founded in 2007 along with a team of computer scientists and drug hunters. Their vision was to change the discovery paradigm through AI – to unlock inaccessible emerging biology and to build a platform capable of learning from all previous data (and mistakes) in the industry. He has led Numerate's efforts to secure multiple rounds of funding, as well as secured several high-profile collaborations with large Pharma companies.

As the company grew, they evolved their business model from being primarily a medicinal chemistry service provider for Pharma to an AI-driven company with both major Pharma alliances and a pipeline of first-in-class therapeutic programs. This has given Guido a unique perspective on the having to simultaneously solve scientific, technological, and business-related challenges of building a successful company based on a proprietary AI platform.

Prior to Numerate, Guido co-founded and was the CTO of another algorithm-focused biotech startup (Pharmix), where he led the development and application of the drug discovery platform.

Guido's background is in Molecular Biology (UC Berkeley) and Bioinformatics (University of Manchester).

P4: Partnering around ADMET Data

Guido Lanza

ADMET data are the ideal data for sharing and generating collective mutually beneficial models and insights due to it's precompetitive nature. We have been handling public and partner data for over ten years and more recently have engaged in a number of data partnerships with larger pharmaceutical companies where they have given us access to large amounts of historical ADMET data. We will discuss our experiences with partner and public data and how these can be improved to produce better machine learning models. In addition, we will explore our recent partnering model that allows multiple pharmaceutical companies to contribute data and receive the benefit of models based on the combined data without exposing their data to each other.

Narayanan Ramaswamy
Global Director Solutions Architecture & Technology Enablement, *Otsuka*

P5

Narayanan Ramaswamy, BSc Physics, BTech Electronic Instrumentation and Computer Science, started his professional career as an Electronic Instrumentation engineer in 1997 at SIEMENS, Mumbai India and after couple of years joined the software industry working with SIEBEL and ORACLE for about 17 years before working for Otsuka. He has worked on variety of industries, projects and in various capacities.

At Otsuka, Narayanan is responsible for Solutions architecture and technology enablement, with focus in the R&D domain.

P5: Data-readiness in a World of AI

Narayanan Ramaswamy, Ashok Upadhyay

Opportunities and Challenges: AI and Machine Learning pose both opportunities and challenges. While these technologies are still emerging, they deliver practical benefits to solve real world problems. To take advantage of these benefits, technical professionals must prepare with the right foundational steps such as developing an AI Strategy, devising data management and data quality processes amongst others. One of the challenges that IT professionals face is data is in silos leading to data fragmentation and difficult to gain actionable insights, data sprawl leads to security and compliance risk and application and services are tightly coupled to data sources preventing modernization.

Solution: Data virtualization[DV] enables information agility. It helps provide a simplified, unified and integrated view of trusted business data in real time or near real time as needed by consuming applications, processes or technologies. DV integrates data from disparate data sources, locations and formats without having to replicate the data. It helps in enabling right time integration and creating data services. DV reduces data sprawl, enables model-driven development, reuse, loose coupling, On Demand access and no batch latency, role-based data access, security and auditing.

Use Case: At Otsuka, data scientists needed data from clinical, financial, project server and ERP. Data is siloed, distributed, and stored in different data stores and some data elements needed to be masked. As an example, for financial ageing AR analysis date ranges are important and so is the financial value, but only for business users and not for technical development team. The legacy system would entail lots of customization to meet the above need. Using DV, we created one single virtual database and access is provided to authenticated users and authorization drives the data visibility. Business users see the visualization developed by developers on real data and developers develop visualizations on masked data. Additionally, this virtual database can now be consumed by data consumers wanting data in any format API, SQL dataset, JSON, XML. In short clinical, financial and ERP data can now be consumed by variety of consumers in variety of formats adhering to InfoSec needs and without having to alter the back-end systems.

Challenges faced: Data virtualization is a relatively newer concept intended to fit within a modern logical enterprise data warehouse to meet the modern analytical needs. There was a bit of learning involved in connecting, composing and in data consumption.

Tim Delisle

Chief Executive Officer, *Datalogue*

P6

Tim Delisle, MSc, is the CEO and co-Founder of Datalogue. Tim's obsession with data stems from too many underslept and over-caffeinated nights working on data problems, inspiring him to co-found Datalogue.

Previously, Tim received his B.S. in Cell Biology and Anatomy from McGill University, his M.Sc. in Computer and Information Science from Cornell Tech, and his M.Sc. in Applied Information Sciences from the Technion—Israel Institute of Technology.

P6: Measuring and Optimizing Time to Data

Tim Delisle

Datalogue is time to data. The company was founded based on the belief that data-driven decisions are the best decisions. The challenge faced by companies and their employees: how to make all created and collected data available for immediate use. 95% or more of data created by organizations is never used - not because it lacks value but because the process to clean and prepare the data has not kept up with the creation of that data.

Datalogue solves that problem. Fast graph abstraction machine learning turns structured and semi structured data - no matter where and how stored - into streams of data. Neural networks ingest the stream and power intelligent classification of each data point - enabling companies to find all HIPAA related data, research data, or business metric even if it is buried in the wrong file, wrong column, or field. Joining disparate data sources with varying schemas (that may change over time) and standardizing the data into a singular output file for any data consumer in an organization enables high velocity time to data.

Datalogue goal: increase the velocity of a decision by increasing the velocity of your time to data. How long it takes to get to usable data, consume the data, and then iterate on the data is what separates the top performing companies.

SESSION IB: PLENARY PRESENTATIONS (cont.):

Session Chair: Christian Baber

Head, R&D IT, *Shire*
Technical Meeting Chair, *PRISME Forum*

Andrew Carroll

Product Lead, *GoogleBrain, GoogleAI*

Andrew Carroll, PhD, is product lead for the GoogleBrain Genomics team, within GoogleAI, where he is responsible for the development of deep learning methods that operate on genomic data and its combination with clinical and imaging data.

Prior to joining Google, Andrew was Chief Scientific Officer at DNAnexus, where he coordinated scientific engagement between DNAnexus and its partners and customers, which included collaborations between GoogleBrain and DNAnexus.

Andrew holds a PhD in Molecular Biology from Stanford University. His PhD research involved comparative genomics in plants. In his postdoctoral work at Lawrence Berkeley Labs he used machine learning approaches to investigate protein localization and modification in mass-spectrometry data.

Identifying and Exploiting Diverse and Difficult Datasets for Training ML

Andrew Carroll

Historically, new software methods have been developed by programmers working with development datasets. New machine learning and deep learning methods present a different paradigm by decoupling the engineering of the development framework from the process of training models. This allows domain experts to leverage existing frameworks and focus on bringing diverse, high-quality data to drive insight.

Here, we present examples – combining deep learning molecular phenotype predictors with cancer sequencing to discover mutations to transcription factor binding sites that may drive tumor evolution. We present how Google's DeepVariant program was improved by identifying diverse and difficult datasets for training.

Mark Musen

Professor, Biomedical Informatics and Biomedical Data Science, *Stanford University*

Mark Musen, MD, PhD, is Professor of Biomedical Informatics and of Biomedical Data Science at Stanford University, where he is Director of the Stanford Center for Biomedical Informatics Research. He conducts research related to open science, metadata for enhanced annotation of scientific data sets, intelligent systems, reusable ontologies, and biomedical decision support. His group developed Protégé, the world's most widely used technology for building and managing terminologies and ontologies.

He is principal investigator of the National Center for Biomedical Ontology, one of the original National Centers for Biomedical Computing created by the U.S. National Institutes of Heath (NIH). He is principal investigator of the Center for Expanded Data Annotation and Retrieval (CEDAR). CEDAR is a center of excellence supported by the NIH Big Data to Knowledge Initiative, with the goal of developing new technology to ease the authoring and management of biomedical experimental metadata.

Dr. Musen directs the World Health Organization Collaborating Center for Classification, Terminology, and Standards at Stanford University, which has developed much of the information infrastructure for the authoring and management of the 11th edition of the International Classification of Diseases (ICD-11).

Dr. Musen was the recipient of the Donald A. B. Lindberg Award for Innovation in Informatics from the American Medical Informatics Association in 2006. He has been elected to the American College of Medical Informatics, the Association of American Physicians, the International Academy of Health Sciences Informatics, and the National Academy of Medicine.

CEDAR: Semantic Technology in Support of Open Science and Improved Knowledge Management

Mark Musen

The past few years has seen a flurry of interest in making scientific data "open" and available online to enable verification of research results and secondary analyses of the data. The buzzword is that scientific datasets must be "FAIR"—findable, accessible, interoperable, and reusable.

The problem is that most scientific datasets are not FAIR. When left to their own devices, scientists do an absolutely terrible job creating the metadata that describe the experimental datasets that make their way in online repositories. The lack of standardization makes it extremely difficult for other investigators to locate relevant datasets, to reanalyze them, and to integrate those datasets with other data.

The Center for Expanded Data Annotation and Retrieval (CEDAR) has the goal of enhancing the authoring of experimental metadata to make online datasets more useful to the scientific community. CEDAR technology includes methods for managing a library of templates for representing metadata, and interoperability with a repository of scientific ontologies that normalize the way in which the templates may be filled out. CEDAR uses a repository of previously authored metadata from which it learns patterns that drive predictive data entry, making it easier for metadata authors to perform their work. Ongoing collaborations with several major research consortia are allowing us to explore how CEDAR may ease access to scientific data sets stored in online repositories and enhance the reuse of the data to drive new discoveries.

SESSION IV: BREAKOUT SESSION

Session Chair: Lars Greiffenberg

Director R&D Information Research, AbbVie Library Sciences & Academic Partnerships

Lars Greiffenberg, PhD, MS, holds a M.S. in Biology and a Ph.D. in Microbiology and has more than 15 years of experience in the field of integrated R&D IT solutions and translational informatics. He held different R&D IT management positions at Aventis Pharma and Sanofi-Aventis in Frankfurt before relocating to the Sanofi site in Toulouse, France where he was Global Head of Solution Center Translational Medicine with responsibility to manage and lead a global program to enable translational science at Sanofi. In 2014 he joined AbbVie in Ludwigshafen (Germany) as director of R&D IT and Translational Informatics. In this role he is heading business IT support covering data and solutions from early discovery up to Medical Affairs. In 2017 he extended his responsibilities including now global Library Sciences at Abbvie. He is driven by the ambition to transform the way we access, consume and leverage literature in the future. He recently established a team at AbbVie, dedicated to use modern methods and algorithms to extract and visualize mechanistic disease information from literature content. In 2018 he further enlarged his area of responsibility to incorporate the Academic Partnerships Organization which is leveraging an AbbVie-Campus at the University of Illinois Urbana-Champaign. Lars is active in several pre-competitive organizations including IMI, PRISME Forum, Pistoia Alliance and EIT-health.

Objective: Create a follow-on paper – from the Fall 2018 paper – on the topic of "Data-readiness in a World of AI"
Methodology: Use the Technical Meeting as a primary information source for the paper.
Use the different breakout groups to examine in details different aspects of the Technical Meeting.
Provide the paper's authors with sufficient detail to write the framework of the paper.

GROUP A - SESSION II: Start-up Pitches
Co-captains: Carol Rohl and Sean Byrnes

Ed Addison (Cloud Pharmaceuticals)
Massimo de Francesco (UCB)
Pete Dhillon (Daiichi-Sankyo)
Joel Ekstrom (Ionis)
Timothy Gardner (Riffyn)
John Overington (Medicines Discovery Catapult)
Filip Pattyn (ONTOFORCE)
Narayanan Ramaswamy (Otsuka)
Jean-Luc Schmidt (Sanofi)
David Sedlock (Takeda)
Etzard Stolte (Roche)

GROUP B - SESSION IA: Plenary Presentations 1 & 2
Co-captains: Nick Brown and Andrew Goldberg

Brandon Allgood (Numerate)
John Apathy (Celgene)
Alastair Binnie (BMS)
Andrew Carroll (GoogleAI)
Dan Chapman (UCB)
M. Hall Gregg (Pfizer)
Jay Krishna (Shire)
Roy Ladd (AbbVie)
Guido Lanza (Numerate)
Natalie Mirutenko (Takeda)
Andy Newsom (CSL Behring)

GROUP C - SESSION III: Poster Presentations 1-3
Co-captains: Martin Romacker and Jake Plummer

Christian Baber (Shire)
Edsel Calliste-David (Astellas)
David Christie (CSL Behring)
Thomas Frei (Novartis)
Klaus Hofenbitzer (Celgene)
Brian Martin (AbbVie)
Tomoyuki Matsunaga (Takeda)
Mark Musen (Stanford University)
Susie Stephens (Pfizer)
Tatsuyuki Takahashi (Mitsubishi Tanabe)

GROUP D - SESSION IB: Poster Presentations 4-6
Co-captains: Jason Tetrault and Andy Beck

Michael Cassidy (Regeneron)
Phil Hajduk (AbbVie)
Hongmei Huang (Genentech)
Bruno Larmurier (Servier)
Francois Midili (Ferring)
Mike Montello (GSK)
Arun Nayar (Amgen)
Michael Shanler (Gartner)
Jon Stevens (AbbVie)
Deep Vaswani (Astellas)

GROUP E - SESSION IB: Plenary Presentations 3 & 4
Co-captains: Jianchao (JC) Yao and Jason Smith

Andrew Allen (Regeneron)
Kelly Caruso (Shire)
Hans Constandt (ONTOFORCE)
John Conway (AstraZeneca)
Tim Delisle (Datalogue)
Martin Erkens (Roche)
Lars Greiffenberg (AbbVie)
Martin Leach (Alexion)
Scott Oloff (Boehringer Ingelheim)
Leonard Sagalov (AbbVie)
Ashok Upadhyay (Otsuka)

SESSION V: MEETING SUMMARY & AWARDS

Session Chair: Christian Baber
Head, R&D IT, *Shire*
Technical Meeting Chair, *PRISME Forum*