PRISME Forum Spring 2021 Technical Meeting

Online, May 19-20, 2021

THEME/BACKGROUNDER: "MACHINE-READY DATA IN SUPPORT OF DATA SCIENCE"

Industry in general, but the biopharmaceutical industry in particular, is increasingly understanding that data is the "new oil".

Data has always been important to the biopharmaceutical industry; data underpinned medicinal chemistry and biology understanding. Data was derived from experimentation carried out simply to observe what happened or indeed to test hypotheses. Those experiments in drug discovery and pre-clinical research, or in non-clinical and clinical development created significant quantities of data that was used to further the biopharmaceutical industry objective of designing and developing new medicines.

However, today, as technology breeds technology, the opportunity to collect data about medicines and the people who use them increases exponentially. New research capabilities create data from increasingly powerfuland varied omics technologies, multiple imaging systems and wearable smart devices. Furthermore, social media plays a significant role. Real-World Data about biopharma therapeutics, obtained in the wider community, increasingly is seen as a valuable tool to understand the effectiveness of a drug in real day-to-day use rather than its efficacy identified in carefully controlled randomized clinical trials.

The industry is seeking to recruit data scientists to help it make sense of this data coming to the industry with volume, variety and velocity to improve the efficiency and effectiveness of biopharma R&D. The data scientists perceive that AI/ML technologies can help bring understanding in this increasingly data-rich environment. However, for these AI/ML tools to be effective, they must have machine-ready access to data.

This PRISME Forum Technical meeting will seek to identify best practices, including management structures, signpost the relevant data standards, identify state-of-the-art tools, and provide exemplar use cases that demonstrate the ROI that can be realized by implementing good data governance.

AGENDA

PRISME Forum Spring 2021 Technical Meeting

"MACHINE-READY DATA IN SUPPORT OF DATA SCIENCE"

WEDNESDAY, May 19, 2021 *(Note: Sessions will be recorded)*	
All timings are stated in US East Coast Time	
09:30	Technical Meeting Introduction **Scott Oloff**, *Boehringer Ingelheim*
09:40	**LECTURE** Putting data Science in the Centre: Observations from EMBL, a Multisite, International Life Science Research Organization **Rolf Apweiler**, *European Bioinformatics Institute*
10:10	*Q&A*
10:20	**SHORT PRESENTATIONS** Introductions **Hongmei Huang**, *Genentech*
10:25	Impowering the Labs by Enabling Data for AI/ML **Jake Lustig**, *Merck*
10:40	EDIS - FAIR and Shared Data Driving PHC **Benjamin Szilagyi**, *Roche*

10:55	Covering Our Data Sets: Making Sense of Data **Mark Borowsky**, *Novartis*
11:10	The Framework for Standardizing Data for Machine Learning **John Apathy**, *BMS* & **Dana Vanderwall**, *BMS/Allotrope Foundation*
11:25	**SHORT PRESENTATIONS - PANEL DISCUSSION** Moderator: **Hongmei Huang**, *Genentech*
12:00	*End of Day 1 Tech Meeting Presentations*

THURSDAY, May 20, 2021 *(Note: Sessions will be recorded)*

09:30	**LECTURE** Leverage Data as Currency to Become a Disruptor **Scott Snyder**, *Breakthru Advisors, LLC*
10:00	Q&A
10:10	**LIGHTNING TALKS** *Introductions* Scott Oloff, *Boehringer Ingelheim*
10:15	From FAIR (Findable, Accessible, Interoperable and Re-usable) to FAIR (Fully AI Ready) **Axel Wilbertz**, *AbbVie* & **Kees van Bochove**, *The Hyve*
10:25	Novartis ID: Building a next-generation life science data intelligence platform with Datagrok **Nik Stiefl**, *NIBR* & **Andrew Skalkin**, *Datagrok)*
10:35	Silo-Breaker **Daniel Huston**, *BMS (bio)* & **Bill Van Etten**, *BioTeam (bio)*
10:45	eSource: Site EHR to Sponsor Using CONFORM Is Now a Reality **Tim Joy**, *Pfizer* & **Munther Baara**, *Edetek*
10:55	**LIGHTNING TALKS - PANEL DISCUSSION** Moderator: Scott Oloff, *Boehringer Ingelheim*
11:25	**Ram C. Iyer**, FDA
11:55	Q&A
12:05	Concluding Remarks **Scott Oloff**, *Boehringer Ingelheim*
12:15	*End of Tech Meeting*

SHORT BIOGRAPHIES

TECHNICAL MEETING CHAIR:

Scott Oloff, Executive Director, IT Research, Development, and Medicine Enablement, *Boehringer Ingelheim*

Scott Oloff, PhD, started his career as a Computational Chemist in Research where he worked for both Biogen and Boehringer Ingelheim.

His primary expertise was building and using machine learning algorithms for potency/ ADMET predictions, docking, and research project support. In 2008 he took a position in Research IT within Boehringer Ingelheim overseeing data analytics and mining technologies.

Over time he has expanded into a number of IT R&D roles where he now oversees all technologies that are shared across the Research, Development, and Medicine business areas (Analytics, Chemistry tools, ELN's, Lab technologies, LIMS, Document Management Tools, etc.).

Scott holds a B.S. in Chemical Engineering from Clemson University and a PhD in Pharmacology from the University of North Carolina.

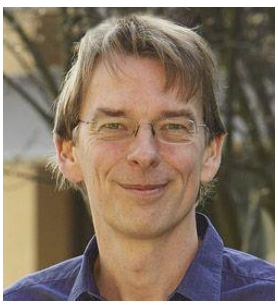
Hongmei Huang, Vice President, Development Sciences Informatics, *Genentech*

As the Vice President of Development Sciences Informatics at Roche Genentech, Hongmei is responsible for the strategic leadership around data management, informatics systems and analytics platforms for translational research and development functions. She is among the key leaders driving the Roche wide effort to make our data Findable, Accessible, Interoperable and Reusable (FAIR).

By connecting science and technology, Hongmei leads organizational drives to transform the data and informatics landscape for the advancement of medicines and healthcare. She is an accomplished scientific and informatics leader with over 25 years of experience in the Pharmaceutical Industry. She started her career as a Research Investigator at Bristol-Myers Squibb and transitioned into Informatics over the course of her career, with leadership roles in various companies including Novartis and Johnson & Johnson. She received her B.S. from Beijing University, M.S. from University of Michigan, and Ph.D. in BioOrganic Chemistry from The Scripps Research Institute.

https://www.linkedin.com/in/hongmei-huang-1081848/

Rolf Apweiler, Director, *European Bioinformatics Institute*



Rolf Apweiler is the Director of EMBL-EBI, together with Ewan Birney. Prior to this position he was Joint Associate Director, after many years of leading protein resources such as UniProt and InterPro. Rolf has made a major contribution to methods for the automatic annotation of proteins, making it possible to add relevant information to proteome sets for entire organisms. He has spearheaded the development of standards for proteomics data, and his teams have maintained major collections of protein identifications from proteomics

experiments (PRIDE) and molecular interactions (IntAct). He also led EMBL-EBI's contribution to the Gene Ontology, was Director of Open Targets, and is now leading the efforts of EMBL-EBI around the European COVID-19 Data Platform.

Rolf received his PhD from the University of Heidelberg in 1994, and has been at EMBL since 1987. His major contribution to the field of proteomics was recognized by the the Human Proteomics Organization's "Distinguished Achievement Award in Proteomics" in 2004 and his election to President of the Human Proteomics Organization, which he held in 2007 and 2008. In 2012, he was elected as a member of EMBO and in 2015 he was elected to an ISCB (International Society for Computational Biology) fellow. Rolf also served over many years on a multitude of Editorial Boards and Scientific Advisory Boards.

https://uk.linkedin.com/in/rolf-apweiler-380044b

Jacob (Jake) Lustig, Associate Vice President, Merck Research Labs, *Merck*

Jake Lustig, Associate Vice President, Merck Research Labs (MRL), leads the Global Clinical IT organization. In this role, Jake is responsible for all of IT for MRL's Clinical Function, inclusive of Global Medical Affairs and Research Quality. In addition, Jake leads the MRL IT Engineering and Project Management Office supporting all of MRL system delivery and operations.

Most recently, Jake led the Merck Manufacturing Quality IT organization focusing on the transformation of the business through digital in the Data Integrity, Quality Management Systems and Manufacturing Labs functional areas. Jake led the Supply Chain IT and MMD Business Insights/Analytics capability. Prior to taking on his responsibilities with Merck Manufacturing IT, Jake worked within the Merck Research Laboratory (MRL) IT organization, leading the Regulatory, Pharmacovigilance and Observational Research IT teams to leverage IT capabilities in the realization of the business strategies.

Prior to joining Merck, Jake was a Senior Manager with Accenture leading large system implementation projects for various life science organizations. In addition, he built and ran several global outsourcing capabilities.

Jake holds a Bachelor of Science degree in Chemical Engineering from the University of Notre Dame.

https://www.linkedin.com/in/jakelustig/

Benjamin Szilagyi, VP, Head Insights Data & Experimental Analytics, *Roche*

Benjamin Szilagyi is Head of Insights Data & Experimental Analysis at F. Hoffmann-La Roche. His function drives PHC focused data and analytics solutions across Pharma Development Data Sciences.

In his past Benjamin has led and contributed to several programs and activities such as the Industry - FDA Collaboration Platform for Data Standards and Exchange, starting up TransCelerate BioPharma Inc and over the last 4 years Enhanced Data & Insights Sharing (EDIS), a Roche R&D program that provides an ecosystem for F.A.I.R. and Shared

biomedical data to advance Precision Medicine through advanced analytics.

https://www.linkedin.com/in/benjamin-szilagyi-2787841/

Mark Borowsky, CIO, *Novartis Institutes for Biomedical Research*

Mark Borowsky is the CIO at the Novartis Institutes for BioMedical Research (NIBR), leading NIBR Informatics. Prior to his current role, Mark led the Information Products and Data Sciences, Informatics Systems, and Scientific Data Analysis teams. NIBR Informatics partners with bench scientists and data science groups embedded in NIBR's disease areas to create software and automation products and services that drive drug discovery.

Before joining NIBR, Mark served as the founding Director of the NextGen sequencing core at Massachusetts General Hospital and Director of Bioinformatics in the Department of Molecular Biology at MGH. He developed both resources into highly collaborative facilities performing next generation sequencing and data analyses in diverse areas of biomedical research. He was a scientist at the Broad Institute, where he contributed to the human genome project and worked on drug resistant *M. tuberculosis* genomes. Mark came to the Broad after working at Incyte, where he managed the microbial pathogen genome database and built an annotation team that identified new splice variants and full length human cDNAs as candidate drug targets.

Mark received a bachelor's degree in Biochemistry and Molecular Biology at Harvard and a PhD in Biology from MIT. Mark did a post-doc in host-pathogen interactions at UC Berkeley.

https://www.linkedin.com/in/borowsky/

Dana Vanderwall, Director, Biology & Preclinical IT, *BMS* and *Allotrope Foundation*

Dana Vanderwall is Director of Biology & Pre-Clinical IT at Bristol-Myers Squibb and is responsible for the strategy to enable data flow, analytics, knowledge capture, and provide general laboratory informatics capabilities to the Biology, Pharmaceutical Candidate Optimization, Drug Safety Evaluation and Veterinary Sciences functions at BMS. He is Chair of the Allotrope Foundation Board of the Directors, contributes projects within the Pistoia Alliance and is Fellow of the Society for Lab Automation and Screening.

Prior to joining BMS Dana was at GlaxoSmthKline and Merck, and integrates experience from work in Computational Chemistry, Cheminformatics, Computational & Structural Biology, and Biochemistry.

Dana received his BS in Biochemistry at the University of Wisconsin, and his PhD in Biochemistry at University of Maryland.

https://www.linkedin.com/in/dana-vanderwall/

Scott Snyder, ***Co-author Goliath's Revenge; President, Breakthru Advisors, LLC***
Senior Fellow, Management Department, The Wharton School, *Adjunct Faculty, The Moore School of Engineering, University of Pennsylvania*

Dr. Snyder is a recognized thought leader in technology and innovation. He has more than 30 years of experience in emerging technologies, business strategy and innovation, and digital transformation for Global 1000 companies and startup ventures. He is the co-author of *Goliath's Revenge: How Established Companies Turn the Tables on Digital Disruptors* and the author of *The New World of Wireless: How to Compete in the 4G Revolution* and numerous other articles on emerging technologies and innovation.

Dr. Snyder is a Senior Fellow in the Management Department at the Wharton School, an Adjunct Faculty Member in the School of Engineering and Applied Science at the University of Pennsylvania, and has lectured at MIT, Babson, Duke, Georgia Tech and INSEAD on Digital Innovation, Decision-making, Business and IT Strategy, Emerging Technologies, Product Design and Development, and Big Data/Analytics. Dr. Snyder is a Digital Economy Project Fellow for the World Economic Forum and sits on the Penn Health Tech Advisory Board. He also holds three patents for personalization engines and has been quoted as a thought leader in numerous publications including CIO Magazine, WIRED, Forbes, Knowledge@Wharton, Los Angeles Times, The Wall Street Journal, CNBC, and the Financial Times.

Dr. Snyder is currently the President of Breakthru Advisors focused on helping enterprises leverage digital and other emerging technologies to accelerate innovation and new venture creation. Dr. Snyder was recently the Global Head of Digital and Innovation at Heidrick Consulting where he is still a Senior Advisor. Before Heidrick, he was the Chief Technology and Innovation Officer at Safeguard Scientifics which provides capital and relevant expertise to fuel the growth of technology-driven businesses in healthcare, financial services and digital media. Prior to Safeguard, Dr. Snyder was the Co-Founder, President and Chief Strategy Officer at Mobiquity (recently acquired by Hexaware), a leader in delivering innovative mobile and digital solutions for enterprises, where he continues to be Chairman of Mobiquity's Advisory Board. Scott helped Mobiquity grow it's healthcare business and helped launch several groundbreaking digital health solutions including the first FDA approved mobile medical app, largest employee wellness app, and flagship patient app for a premier US Health System. He has consulted on digital and business model innovation in the healthcare space for companies such as GSK, IQVIA, Novartis, J&J, Alcon, Medtronic, UPMC, Vitality, Weight Watchers, Guidewell, and Humana.

Dr. Snyder has held executive positions with several Fortune 500 companies including GE, Martin Marietta, and Lockheed Martin, has been the CEO of a leading strategic planning firm, Decision Strategies International (now part of Heidrick & Struggles), and has also started business ventures in software including OmniChoice, a CRM/Analytics applications provider. He also serves on the Board of Directors for Fulton Financial Corporation (FFC).

Dr. Snyder earned his BS, MS, and PhD in Systems Engineering from University of Pennsylvania.

https://www.linkedin.com/in/scottsnyder5g/

Nikolaus (Nik) Stiefl, *Associate Director Data Science, GDC CADD, NIBR*

Nik Stiefl is a data scientist with a passion for data and early drug discovery. Trained as a pharmacist, Nik switched to chemoinformatics, machine learning and molecular modelling early in his career.

As part of multiple small-molecule drug discovery projects he delivered a range of discovery candidates. He also developed new scientific algorithms as well as software for the community.

Since 2019, Nik is the product owner of two major digital initiatives at NIBR – one more focusing on machine learning and the other on data science and visualization.

https://www.linkedin.com/in/nikolaus-stiefl-39583b25/

Andrew Skalkin, CEO, *Datagrok*

Andrew Skalkin is a computer scientist, entrepreneur, CEO and principal architect at Datagrok, a company behind the next-generation data platform for analyzing complex data commonly encountered in life sciences.

Prior to founding Datagrok, Andrew has had a wide exposure to biopharma. During his 17-year long stint with Janssen, he held multiple roles (software engineer, architect, scientist, manager) while developing and supporting novel solutions for multiple domains such as cheminformatics, health informatics, LIMS, bioinformatics, biomarker research, drug development and manufacturing, statistical process control, biosensor data integration, and clinical data analysis. Notable work includes ABCD (an integrated drug discovery informatics platform used daily by thousands of scientists), and JAKE (Janssen Autism Knowledge Engine - a biosensor-driven platform for enabling clinical trials for ASD).

Datagrok is a next-generation platform designed from scratch to understand scientific data.
It connects people, data, and algorithms together in a novel way, and uses a number of technological breakthroughs to enable highly-interactive analysis of big datasets right in the browser.

Out of the box, it provides a compelling set of enterprise-ready capabilities for data discovery and access, data governance and transformations, exploratory data analysis, advanced visualizations, ML & AI, integration with R and Python, data augmentation, and collaboration. The platform is highly extensible, and leverages an open-source ecosystem of extensions, including first-class support for cheminformatics, bioinformatics, and a repository of scientific methods.

Datagrok's mission is to help people understand data on a fundamentally higher level.
While the platform is domain-agnostic, the company's initial focus is in life science and biopharma. Our customers are some of the world's biggest pharmaceutical companies that use Datagrok in multiple contexts, from early drug design and discovery to clinical data analysis.

https://www.linkedin.com/in/andrew-skalkin/

Axel Wilbertz, PhD Student at AbbVie/Heidelberg University, Biologics Formulation Department & Data Science, AbbVie

Axel Wilbertz joined AbbVie in 2015. After his masters degree at AbbVie, he worked as a data engineer/data scientist for three years at the biologics formulation department. He evolved the internal data strategy to efficiently process and standardize protein stability data of pipeline molecules, which laid the first step towards FAIR data. He started his PhD in 2018 at the Medical Faculty Mannheim, Heidelberg University with the goal to show that the application of FAIR on complex analytical methods (liquid chromatography) is beneficial for the application of predictive analytics.

Axel is keen to develop a data driven organization and to push forward the digital conversion at AbbVie. His main interest is a company-wide data strategy and a practical FAIR implementation for pharmaceutical data.

From FAIR (Findable, Accessible, Interoperable and Re-usable) to FAIR (Fully AI Ready)

- Overcoming the gap between the current FAIR guiding principles and achieving machine actionability.

The FAIR guiding principles aim to make digital resources more Findable, Accessible, Interoperable and Re-usable (FAIR) by utilizing adequate metadata.

In theory the FAIR principles should provide an ideal starting point for using data for advanced data science such as machine learning. Nevertheless, no successful practical FAIR implementation was published, demonstrating the use of FAIR to enable artificial intelligence. Is there a reason for that?

This problem may originate from the current focus on making data Findable and Accessible, rendering data machine readable, but are insufficient for application of AI algorithms.

What is currently missing is a comparability functionality to incorporate SME domain knowledge to distinguish apples from oranges to achieve true interoperability and reusability (I and R) by an algorithm (machine actionability). Besides, the FAIR guiding principles hardly cover data quality as well as scientific comparability aspects. Therefore, a combination of a metadata-based semantic model (open ontologies including RDFs) with a comparability feature implemented as a decision tree reflecting SME knowledge is proposed. The combination renders biologics data not only machine readable, but fully machine actionable as a prerequisite to achieve AI-readiness of data. A case study will demonstrate, how such an approach has been implemented in the Biologics formulation development at AbbVie.

All authors are employees of AbbVie and may own AbbVie stock. AbbVie sponsored and funded the study; contributed to the design; participated in the collection, analysis, and interpretation of data, and in writing, reviewing, and approval of the final publication.

https://www.linkedin.com/in/axel-wilbertz-82a61a183/

Kees van Bochove, Founder, *The Hyve*

Kees van Bochove is founder of The Hyve, a company dedicated to the support and facilitation of open source, open standards and open data in biomedical informatics. Kees is active in many open source biomedical software development communities such as i2b2/tranSMART, cBioPortal, OHDSI, RADAR and GA4GH, and in precompetitive projects such as IMI EHDEN, PIONEER, BigData@Heart, H2O, FAIRplus, GO-FAIR, Pistoia Alliance, DTL, tranSMART Foundation, Movember Foundation and Digital Biomarkers. He has been involved with the FAIR movement since the initial Lorentz workshop in 2014, and was one of the initiators of the FAIR implementation working group in the Pistoia Alliance.

Today, Kees' main expertise and engagements are as Principal Consultant, advising pharma companies, academic hospitals as well as patient and health data networks on their FAIR Data Strategy and advising and leading implementation projects with teams from The Hyve.

In this role, he has assisted several top 20 pharmaceutical companies build their R&D data strategy, hospital network CDOs and CIOs build translational medicine infrastructures, and national governments and programs such as the Dutch HealthRI, the Swiss Personalized Health Network, and the German Medical Informatics Initiative to build out their health data network infrastructures.

Read more on Kees' view on data governance in his blog: https://blog.thehyve.nl/blog/fair-is-like-a-fractal

https://www.linkedin.com/in/keesvb/

Tim Joy, Head of GPD Clinical Operations Strategic Partnerships, *Pfizer*

Tim Joy is responsible for leading a team tasked to identify and develop creative, scalable, technical and process related solutions for unmet global clinical trial needs.

Tim has been at Pfizer since 2003, when he joined Development Informatics. In Development Informatics Tim was responsible for implementation and support of Pfizer Electronic Data Capture systems. Since joining Global Product Development in 2006 Tim has held multiple positions with increasing responsibility. Tim represented Pfizer on an Industry council to redesign and implement enhanced versions of Electronic Data Capture. Additional key programs of work include the development and implementation of a global Clinical Trial data warehouse.

Tim has been involved in numerous projects and initiatives to redesign Investigator Site Support, implementation and usage of novel Patient Technologies, and creation of Pfizer's eSource strategy.

Before joining Global Product Development Tim was head of Pfizer's Global Clinical Data Services Technical Operations group.

https://www.linkedin.com/in/tim-joy-56a1455/

Munther Baara, VP Product Strategy and Innovation, *EDETEK*

Munther Baara is the product strategy and innovation lead as well as EDETEK's major industry partnerships. Munther has 25 years of experience in leading the development of numerous clinical systems and business solutions. He has held roles of increased seniority in biotechnology and in the world's largest pharmaceutical companies, leading global teams and technology implementations. Munther's experience has been focusing on innovation, trends and emerging technologies in support of the paradigm shift in the execution of clinical trials.

Munther was most recently the head of New Clinical Paradigm at Pfizer where he led and/or contributed to transformative technological breakthroughs. Munther also worked on mClinical patient journey (mobile and sensor), blockchain, returning data to patients, Shared Investigator Platform (SIP), clinical aggregation layers, eSource (Industry-Pioneering EHR Data Transfer) and TransCelerate BioPharma Projects. Also, he was awarded the 2018's Top 20 Innovators in clinical trial advancements by CenterWatch.

https://www.linkedin.com/in/mbaara/

Ram C Iyer, Chief Data Officer, *FDA*

As the Chief Data Officer of the FDA, **Ram C Iyer** has the accountability to develop and execute an agency wide data modernization strategy, building robust central functions that can be leveraged by the centers and the agency for high value decisions. The scope spans the entire stack from data identification to actionable decision, including data policies and governance.

Ram is an industry and peer recognized data and technology professional with experience in the Pharma, Consulting, Telecom and International Government organizations. His expertise includes Data and Decision Sciences, Digital and Technology Architecture, and Talent Development with a focus on building collaborative partnerships and Ecosystems.

Before joining the FDA, Ram was the Head of Enterprise Architecture and Executive Director of Analytics Center of Excellence at Bristol Myers Squibb (BMS). He helped jumpstart several Data and Analytic practices at BMS including enterprise class platforms for reproducible research, model management and visual analytics. He also built a thriving network of data scientists, data analysts, visual story tellers, and agile specialists to tackle urgent and complex problems in the organization.

Ram received his Masters Degree in Computer Science from the New Jersey Institute of Technology and Bachelors Degree in Mathematics from the University of Madras, India. He is also trained in several complementary skills such as Enterprise Analytics, System Dynamics and Design Thinking from leading institutions in the US.

https://www.linkedin.com/in/ram-c-iyer-271267/