Allotrope® Connect Day 1: Wednesday, 15-May-2024

10:00-10:15 Welcome (Frederick Francois Chesneau, BASF and Janet Cheetham, Allotrope Foundation)

- **Frederick Francois Chesneau;** Head of Laboratory Data Solutions - Digital Laboratory, BASF Digital Solutions.
- **Janet Cheetham**; Chair of Allotrope Foundation.

12:15-12:30 Lightning Talks: Member Updates on Digital Journeys: Amgen (Janet Cheetham, Amgen)

Amgen: Updates (Janet Cheetham, Amgen)

- **Janet Cheetham;** Amgen

12:30-13:15 Keynote presentation by Stefan Dreher, BASF (Stefan Dreher, BASF)

- **Stefan Dreher**; Senior Vice President Digitalization of R&D at BASF.

14:00-14:30 Lightning Talks: Member Updates on Digital Journeys: Amgen, Johnson & Johnson, MSD (Gang Xue, Johnson & Johnson. Vincent Antonucci, MSD)

Johnson & Johnson: Secret Sauce to Structured Data & Enable Digital Pharma 4.0 (Gang Xue, Johnson & Johnson)

Throughout years' of cross-functional collaboration, research and development of each pharmaceutical candidate accumulates tremendous amount of data. However, due to the segregation and heterogeneity of these data, knowledge extraction has been tedious and limited while dossier authoring becomes excruciating exercise. Enterprise data lake and ontology enabled data curation and semantic transformation enables Structured Content and Data Management that democratizes scientific data, enabling deep data analytics and automated dossier authoring.

- **Gang Xue;** Sr. Scientific Director at Johnson & Johnson.

  Gang Xue, Sr. Director at Johnson & Johnson, leading Data Integration Modeling within Therapeutic Development & Supply, with prior working experience at Amgen and Pfizer. Gang is one of the co-founders of Allotrope and a current EC member.

MSD: Allotrope Implementation Update (Vincent Antonucci, MSD)

MSD is a founding member of Allotrope Foundation and has been part of the pre-competitive team developing and refining the Allotrope technology framework for several years. This brief presentation is an update on MSD's individual journey implementing AFO / ASM as a key enabler to realizing our organizational goal of frictionless FAIR data flowing from instruments to insights for scientists.

- **Vincent Antonucci;** Technology Product Line Lead at MSD.

Vinny has over 30 years experience at MSD as a scientific, regulatory, and technology product leader and with multiple pre-competitive consortia, all with a single goal to foster collaboration across interfaces that help drive science. Vinny is also the current Vice Chair of Allotrope Foundation.

14:30-15:00 Leveraging knowledge graphs for efficient use of laboratory instrument data (Jindrich Mynarz, Maksym Skoryk, MSD)

Since Allotrope standards and knowledge graphs are both based on the semantic web standards, Allotrope Simple Models can be used as a data exchange format for construction of knowledge graphs. We argue that knowledge graphs provide a powerful leverage to reduce the friction involved in combining laboratory instrument data from heterogeneous data sources and make the data analysis-ready. Moreover, this is a generic and scalable approach not specific to any instrument make or instrument type.

In this presentation, we demonstrate the value of connecting laboratory instrument data in knowledge graphs. In particular, we focus on queries across multiple laboratory instrument runs, coming from different instrument makes or instrument types, such as when aggregating measurements for a given sample. This is feasible thanks to the standard Allotrope Foundation Ontologies describing the data in a uniform way and thanks to graph databases that enable these kinds of joins. We cover semantic queries (e.g., query expansion via ontological relations), queries for self-describing data (e.g., getting definitions of ontology terms), highly expressive queries (e.g., validation of department-specific data quality rules), queries tracing the origin of data (e.g., for audit control), and federated queries (e.g., enrichment with contextual data).

We share practical considerations, reusable patterns, and implementation experience from constructing a knowledge graph out of laboratory instrument data from ASM files at MSD. We conclude with a future appeal for reducing the overhead with copying data by standardizing it directly at the point of its capture by laboratory instruments, and making it thus ready for efficient and immediate use via knowledge graphs.

- **Jindřich Mynarz**; Ph.D. Research Data Engineering lead, RaDS IT. MSD, Prague, Czech Republic.

  Jindrich Mynarz leads a data engineering team supporting pharmaceutical research at Merck Sharp & Dohme. He has 15 years of experience with semantic web technologies both from academia and industry. He is an active user and contributor to Allotrope standards.

- **Maksym Skoryk**; software engineer. MSD, Prague. Czech Republic.

Maxim is a software engineer with a background in chemical engineering. He is experienced in building scalable data pipelines, cloud infrastructure, and analytical solutions for life sciences applications. His passion is to help organizations efficiently use their data to accelerate research and development of better healthcare solutions

15:00-15:30 Standardizing Data, Amplifying Insights: ZONTAL's Methodology for Allotrope-Powered Generative AI (Dennis Della Corte, Zontal)

In the era of Generative AI, harnessing the power of standardized data has become an imperative for unlocking unparalleled insights and driving innovation. With the rapid advancements in Generative AI, the need for

structured and standardized data has never been more crucial. The ZONTAL platform offers a cutting-edge solution to digitalize methods in the Allotrope Simple Method. This presentation puts a special focus on USP, Pistoia, and Empower methods. By seamlessly integrating with Allotrope's comprehensive data models, ZONTAL prepares organizations for the transformative potential of Generative AI applications.

This presentation will delve into the intricacies of the ZONTAL platform, showcasing its capabilities in streamlining method digitalization and enabling seamless data integration within the Allotrope framework. Attendees will gain valuable insights into how ZONTAL positions organizations at the forefront of data standardization, unlocking new avenues for innovation and collaboration in the era of Generative AI. As we stand on the precipice of a data-driven revolution, the synergy between platforms like ZONTAL and Allotrope's standardization efforts paves the way for a future where scientific discoveries and breakthroughs are no longer limited by data silos, but instead, propelled by the boundless potential of unified, structured information.

- **Dennis Della Corte**; Chief Science Officer, Zontal.

  Prof. Dr. Dennis Della Corte serves as Chief Science Officer at ZONTAL and Director of the Consortium of Molecular Design at BYU. His research applies artificial intelligence and advanced analytics in the fields of protein engineering, drug discovery, augmented pathology, and laboratory automation.

15:45-16:15 OpenChrom meets ADF and ASM (Matthias Mailänder, Lablicate)

OpenChrom is a platform independent multivendor chromatography data system with additional support for spectroscopy and molecular biology. All existing file format plugins can now convert from and to ADF/ASM. A new visual ADF explorer allows inspecting and querying data.

- **Matthias Mailänder;** Software Engineer at Lablicate GmbH.

  Matthias Mailänder is a food chemist who works as a software engineer at Lablicate GmbH in Hamburg working on OpenChrom a multi-vendor chromatography data system. He is a strong advocate of Open Source software and FAIR data.

16:15-16:30 Addressing Implementation Challenges of Instrument Data Converters at IFPEN (Maxime Visconte, IFPEN)

IFPEN's initiative to develop an open-source instrument converters library to the Allotrope Simple Model (ASM) holds promise for advancing data interoperability in the scientific community. However, the journey towards standardization is not without hurdles. IFPEN confronts the quintessential challenge in data science: ensuring data quality. For example, the diversity in naming conventions for samples and their associations with Laboratory Information Management System (LIMS) references presents an obstacle, necessitating the creation of bespoke converters for each laboratory. The discussion explores IFPEN's strategy to overcome these challenges and highlights the importance of meticulous data management practices in ensuring the success of standardization efforts.

- **Maxime Visconte**; Director of IT for IFPEN, France's leading research institute on the energy transition.

After a M.Eng in Chemical Engineering, and a few years in the industry working as an Advanced Process Control engineer, Maxime is now in charge of facilitating data acquisition across all labs and pilot-scale industrial units for IFPEN researchers.

16:3O-17:00 Standardizing and unifying unstructured instrument data with Large Language Models (LLMs) (Andrew Chen)

Most laboratory instrument data is unstructured and unsuitable for downstream analysis, machine learning, and compliance applications. Converting this data into useful formats like ASM is a critical requirement for any data-driven R&D organization. However, this has so far been an intractable problem — there are too many instrument types, formats, and standards.

Here we introduce a new, nuanced application of large language models (LLMs) that can rapidly generate highly accurate and scalable converters from instrument data to standardized formats. We introduce three components: test set generation with domain knowledge, attention tracing for scalable human validation, and code synthesis under an agentic loop. The approach has been validated on a small open-source instrument data set, although further evaluation is limited by our ability to acquire more raw instrument data.

This talk will also serve as "a busy person's intro to LLMs in the lab." What they are, their most impressive and surprising capabilities, and where the hype falls short, using concrete examples from our work.

- **Andrew Chen**

  Andrew Chen is a former Stanford Computer Science (MS), Stanford Biology (BS), and Y Combinator. More at awchen.com

17:00-17:15 Allotrope Foundation's Branding and Marketing Refresh Project (Kashef Qaadri, Allotrope Foundation Partner Network Lead)

- **Kashef Qaadri;** Allotrope Foundation Partner Network Lead

17:00-17:15 Closing remarks (Frederick Francois Chesneau, BASF and Vinny Antonucci, Allotrope Foundation)

- **Frederick Francois Chesneau;** Head of Laboratory Data Solutions - Digital Laboratory, BASF Digital Solutions.
- **Vinny Antonucci**; Vice Chair Allotrope Foundation.

Allotrope® Connect Day 2: Thursday, 16-May-2024

9:00-9:15 Opening remarks (Frederick Francois Chesneau, BASF and Vinny Antonucci, Allotrope Foundation)

- **Frederick Francois Chesneau;** Head of Laboratory Data Solutions - Digital Laboratory, BASF Digital Solutions.
- **Vinny Antonucci**; Vice Chair Allotrope Foundation.

9:15-9:45 Unleashing Lab Connectivity: LADS and ASM Revolutionizing Seamless Lab Integration and Data Flow (Heiko Fessenmayr, Agilent Technologies)

There's a long dreamed dream of a combining standardized lab instrument communication and standardized lab data formats for interoperability in the digital lab. This is a very exciting, forward-looking presentation and demonstration to show how this dream can come true using OPC-UA based Laboratory Agnostic Device Standard (LADS) for instrumentation integration and Allotrope Simple Model (ASM) for data accessibility. It shows how LADS and ASM (exposed as the OPC-UA Information Modeling layer) could play nicely together to make the perfect couple allowing interoperability across diverse laboratory devices. Allotrope Simple Model (ASM) is a JSON(-schema) based standard. Elevating ASM to an OPC-UA information model and integrating into a LADS driver offers new possibilities in multiple dimensions.

A proof of concept with the Liquid Chromatography ASM as OPC-UA Information Model will be demonstrated. Whether you are a laboratory hardware/software provider intending to adopt LADS and ASM, a lab integrator or data scientist, this presentation based on OPC-UA LADS and the Allotrope ASM shows the future for streamlined laboratory workflows, instrument integration, and improved data accessibility.

- **Heiko Fessenmayr**; Software R&D LabInformatics System Architect, Agilent Technologies.

  Heiko Fessenmayr is a distinguished figure in the realm of analytical laboratory standards, focusing on the development of data formats and communication protocols. With over five years of dedicated service, he represents Agilent Technologies at the Allotrope Foundation, contributing significantly to the advancement of the ADF/ASM standard data format.

  In addition to his foundational work, Heiko plays an instrumental role in enhancing analytical laboratory IoT communication standards through OPC-UA technologies. His leadership as chairman of the OPC-UA ASI/CAISI working group is pivotal in the standardized integration of analytical devices into comprehensive data systems.

  Heiko's expertise extends to the conceptualization of LabInformatics solutions, particularly for chromatography data systems and analytical data content management systems. His comprehensive experience, spanning over two decades with HP/Agilent, encompasses software engineering, project and program management, technical marketing management, and, since 2015, his influential role as an R&D LabInformatics System Architect.

9:45-10:15 Using ChatGPT for automatic asset onboarding for Industrial IoT projects (Erich Barnstedt, Microsoft) – Remote presentation

Asset onboarding still takes the longest time in an Industrial IoT project! Can we use generative AI technology to automate the asset modelling. Also, can we standardize the configuration interface for popular industrial connectivity solutions for better integration and what roles do open standards play in the above?

- **Erich Barnstedt**; Chief Architect Standards, Consortia & Industrial IoT, Azure Edge + Platforms.

  Erich has worked in various engineering roles at Microsoft for over 21 years, initially in the Windows team and later in the Azure team. Throughout his career, he worked in the automotive and manufacturing verticals and is the founder of both the Windows and the Azure Industrial IoT teams and is the inventor of many Industrial IoT products. More recently, he shifted his work to the support of open standards in Microsoft products as well as commitments to open-source and consortia work in the Azure Edge & Platform team at Microsoft as Chief Architect. He is the holder of various IoT-related patents and has a bachelor and two master's degrees in computer science from Trinity College, Dublin.

10:15-10:45 Unlocking the automation potential via open standards (Kathrin Cohen and Tobias Umbach, BASF)

In today's data driven world, corporate R&D units are constantly seeking ways to unlock the potential of automation to enhance efficiency and productivity. Open data standards have emerged as a powerful tool in this endeavor, facilitating integration and interoperability across various systems. This presentation aims to describe BASF's motivation for adoption of the Allotrope simple model, concrete use cases, as well as open challenges.

- **Kathrin Cohen;** Head of Lab Automation Solutions, BASF.

  Kathrin Cohen currently leads the Group of Digital Lab Process Design in Group Research at BASF. With over 12 years of industry experience, she combines chemical innovation with digital transformation. Kathrin holds a PhD in Chemistry and an undergraduate degree in Chemical Engineering. Throughout her career at BASF, she has worked across various R&D fields, from product development for oil recovery to digitalization and automation for dispersions and resins.

- **Tobias Umbach**; Automation Research Specialist
  Tobia Umbach is an experienced professional with a background in experimental physics. He started as a lab team leader for atomic force microscopy at BASF and currently develops lab data management solutions for R&D laboratories.

11:00-11:30 True Instrument Integration Requires More Standardization (Sven Arenz, BASF)

Instrument integration is crucial for the process of data exchange between LIMS and measuring device, it reduces the risk of errors and data gaps, and allows a high degree of automation in instrument usage. In this presentation I would like to talk about our evolution in instrument connection and why standardization is key to success.

- **Sven Arenz**; Digitalization Research Scientist at Analytical & Material Science, BASF.

Sven Arenz holds a PhD in Chemistry and works currently as a Digitalization Research Scientist in Material physics and Analytics at BASF. Throughout his 25-year career as researcher in industry and academia, he has been driving the use of digital technologies in many laboratories. At BASF he is developing and implementing new digital tools and techniques to streamline laboratory workflows, reduce errors, and improve data quality.

11:30-12:00 Archiving of electronic raw data in the GxP environment - the ADF data format (Burkhard Matthes, Boehringer Ingelheim) – Remote presentation

This presentation describes the use of the Allotrope data format in the laboratory environment under GxP regulations. A validated concept for converting and archiving electronic raw data from Agilent systems into ADF format is presented.

- **Burkhard Matthes,** Head of IPT at Boehringer Ingelheim Pharma GmbH & Co. KG. Department Development of New Chemical Entities.

  Dr. Burkhard Matthes studied chemistry at the University of Bayreuth. He then pursued a Ph.D. in Organic Chemistry, focusing on the synthesis of organic liquid crystals and quantum mechanical calculations of molecular structures. Following his Ph.D., Burkhard worked as a Postdoctoral Researcher in the Research Department at Boehringer Ingelheim. He then transitioned to the Lab of Combinatorial Chemistry at Boehringer Ingelheim, where he worked on chiral separations. Since 2011, Burkhard has been the Head of In Process Testing (IPT) in the Department of Development of New Chemical Entities at Boehringer Ingelheim. Current main interests: analytical data formats, GxP regulations and requirements regarding data integrity, and integration of laboratory data in cloud systems.

12:00-12:30 Allotrope Working Groups and Release Update (Ben Woolford-Lim, Allotrope Product Team)

This update presents the latest activity and developments across the Allotrope Working Groups (Chromatography, Modeling, Mass Spectrometry, Plate Reader, and Flow Cytometry), highlights key releases, the supporting DevOps infrastructure to streamline the ontology and development lifecycle, and provides updates on product enhancements.

- **Ben Woolford-Lim;** Senior Developer at Allotrope Foundation, Allotrope Product Team.

  Ben is a technical expert in all things Allotrope, from the original Allotrope Data Format APIs, to the Allotrope Foundation Ontologies, and the current generation of Allotrope Simple Model schemas and patterns. He joined the Foundation as Senior Developer in 2020, working with Allotrope prior to that as a GSK representative since 2016. His expertise is in software engineering and semantics, with a background in maths, physics, and computer science.

13:15-13:40 AFO + IDMP-O: Connecting pharma data across the entire value chain with a network of interoperable industry ontologies (Heiner Oberkampf, Accurids & Pistoia Alliance)

Improving the data foundation across departments in large organizations and the industry requires not only building semantic data standards like Allotrope and IDMP-O. To reap the envisioned benefits of those standards, we must connect them and work towards broad end-to-end adoption. The talk will provide a brief overview of the ISO IDMP Ontology and describe how a network of ontologies with AFO, IDMP-O, and others helps to realize synergies from different standardization efforts to deliver on their promise with concrete pharma use cases.

- **Heiner Oberkampf;** CEO & Co-Founder ACCURIDS.

  Dr. Heiner Oberkampf is the CEO & Co-Founder of ACCURIDS. Heiner has worked with pharma and life-science organizations over the past 10 years on various data governance and integration topics. His focus is to speed up the adoption of public and enterprise data standards to transition towards a data and analytics-centric approach to organizing and using information. Heiner has contributed to data standardization efforts in Allotrope and is co-leading the ISO IDMP Ontology initiative.

13:40-14:00 IDMP Ontology - Bridging Pharma Efficiency (Gerd R. Kleemann, Amgen & Pistoia Alliance)

The presentation discusses how to implement an external ontology - factors & scenarios.

- **Gerd R. Kleemann;** Director Data Sciences, Operations Data Strategy & Governance in the Department of Operations Transformation & Data Strategy at Amgen Inc.

  In his work, Gerd is driving digital transformation across the Operations Value Chain through the continuous delivery of valuable, high-quality, and durable data assets that are interconnected and expandable.

14:00-14:30 Pharmaceutical CMC ISA-88 Process Ontology (Cameron Gibbs, Crown Point & Pistoia Alliance)

The pharmaceutical CMC process ontology is based on the ISA88/95 framework to standardize laboratory and plant production process recipes to establish standardized definitions, facilitate digital technology transfers, and integration with execution systems in order to capture structured process data for material lot genealogy tracking, streamlined technology transfers, and advanced process analytics, thereby enhancing efficiency and transparency throughout the pharmaceutical production lifecycle. This talk will introduce the ontology, including its current status, relationship with AFO, and sample data applications.

- **Cameron Gibbs;** Knowledge Engineer, Crown Point Technologies.

  Cameron Gibbs is a Knowledge Engineer with Crown Point Technologies, with a background in both industry and academia building and implementing ontologies and knowledge graphs in many domains including for Pharmaceutical and Life Sciences use cases.

15:00-15:30 Natural Language Processing-Based Extension of the Allotrope Foundation Ontology (Alexander Behr, TU Dortmund University) – Remote presentation

In this talk, a brief introduction into the extension of the Allotrope Foundation Ontology (AFO) by new concepts based on scientific literature is presented. Natural Language Processing (NLP)-based concept extraction is used to extend an ontology with new classes and relations. With this the generation of knowledge graphs through NLP-assisted text-mining is presented, which enables reuse of AFO entities and the structured querying of the extracted knowledge.

- **Alexander S. Behr;** PhD-Student at the Laboratory of Equipment Design, TU Dortmund University.

  M.Sc. Alexander S. Behr studied chemical engineering at TU Dortmund University, where he also received his master's degree in 2020. Since 2020 he is working as a PhD-Student at the Laboratory of Equipment Design in context of the NFDI4Cat Project. His research focuses on the development of automated workflows for ontologies in the domain of catalytic sciences and process engineering.

15:30 16:00 ASM Converter Implementation: Strategies & Lessons Learned (Joe Negri, Benchling) - Remote presentation

"allotropy" is an open source Python library developed by Benchling, designed to convert instrument data into the Allotrope Simple Model (ASM). The library's goal is to seamlessly process text or Excel-based instrument software output and generate a JSON representation that adheres to the published ASM schema. JSON is chosen for its readability and compatibility with most modern systems, requiring no specialized tools for human verification. Join us for a presentation where we'll discuss the implementation strategies and lessons learned while developing this library.

- **Joe Negri;** Product Manager for Automation and Analytics at Benchling.

  Joe Negri is a Product Manager for Automation and Analytics at Benchling. Joe worked as a research scientist for over 15 years before transitioning into product development. Prior to joining Benchling, Joe was a Product Manager at Agilent Technologies, and previously worked at the Broad Institute of Harvard and MIT, and the Dana-Farber Cancer Institute. Joe received his PhD in Biological and Biomedical Sciences from Harvard University, and bachelor's degree in neuroscience from Trinity College Dublin.

16:00-16:30 Hackathon Readout (Ben Woolford-Lim, Allotrope Product Team)

Hackathon Readout.

- **Ben Woolford-Lim;** Senior Developer at Allotrope Foundation, Allotrope Product Team.

16:30-17:00 - Reflection: key themes from the meeting / potential next steps (Frederick Francois Chesneau, BASF and Janet Cheetham, Allotrope Foundation)

- **Frederick Francois Chesneau;** Head of Laboratory Data Solutions - Digital Laboratory, BASF Digital Solutions.
- **Janet Cheetham**, Chair of Allotrope Foundation.