



Measurement, Information and Innovation: Digital Disruption in the Chemical Sciences

Abstracts

Session 1: Data acquisition

The purpose, practicalities, pitfalls and policies of managing and sharing data in the UK. Dr Danny Kingsley, Head of Scholarly Communication, Cambridge University.

Why do we share data? What is involved? And what should we look out for? This talk will be a brief overview of aspects of managing and sharing data beginning with the latest developments in UK data management including the policy positions of institutions, funders, government and publishers. Considering the basics of research data management, there will be an exploration of the skillsets required for managing and curating data, and the current opportunities (or otherwise) for training in this area. The practice of sharing data requires the work to be intelligible, assessable and reusable which comes down to the metadata and takes us back to research data management. When preparing to share data, issues such as who owns the data arise - there are potential concerns about collaborations when one party is required to share data and the other is not. The talk will conclude with a brief look at the consequences of not sharing data, and how data is discovered and shared through institutional, cross disciplinary and disciplinary repositories and new data journals.

Allotrope Foundation, how a data standard and taxonomy for analytical data will support data integration and knowledge generation. Dr Gerhard Noelken, Technology and Innovation Group, Pfizer Pharmaceutical Sciences.

This presentation will provide an update on the progress of the Allotrope Foundation towards delivering a Framework solution for managing analytical chemistry data throughout its lifecycle. It will include details on the deliverables, timelines, and results not only on the Allotrope Framework delivery but also on the Integration projects happening across many of the Allotrope member companies.

The intelligent analytical laboratory is moving closer to reality; an automated laboratory where data, methods and processes are seamlessly shared between software applications and analytical instruments.

The Framework leverages emerging standards including: the Allotrope Data Format to store the analytical chemistry results, a set of taxonomies to provide accurate metadata for laboratory systems; and class libraries to provide access to data, metadata, and business objects.

Why we should embrace learnings from other industries. Peter Boogaard, Industrial Lab Automation.

Why can we connect almost any smartphone in our personal lives and why do we need a professor to transfer simple results in our scientific laboratories to our lab computer systems? What can we learn how other industries with comparable scientific, regulated and safety constraints? Cheating is allowed, to re-align our scientific industry and to take advantage of learnings from other industries. This presentation will be highlighting some concrete examples and ideas to start the mindset shift.

Session 2: Data and Information Management

Automation in Analytical Laboratories: A teaching perspective. Professor Krishna Persaud, School of Chemical Engineering and Analytical Science, University of Manchester.

Black box measurement devices and multiuser information management systems are causing glassware and laboratory notebooks to become increasingly rare in the analytical laboratory. While the throughput of these laboratories becomes very high, a threat that is often articulated is the dehumanisation of laboratory activities, reducing the demand for intellectual input, and negation of any fundamental understanding of the science and technological processes in use. The convergence of laboratory automation and laboratory informatics means that traditional ways of teaching analytical science students have to be thrown out of the window and new strategies developed so that the students are taught basic scientific understanding of the principles involved in measurements, but also become familiar with laboratory information management systems and new techniques of knowledge management. This presentation will reflect on potential best practices for teaching undergraduate and graduate students analytical science in the context of laboratory automation.

HELM (Hierarchical Editing Language for Macromolecules): A Pistoia Alliance project providing a standard for the storage and exchange of macromolecule structural information in the information ecosystem. John Wise, Executive Director Business Development - Europe, Pistoia Alliance.

Life science R&D, and in particular pharmaceutical industry R&D, is experiencing an unprecedented and perhaps inexorable trend towards externalization and virtualization. As such efficient information flows through the information ecosystem are self-evidently essential to support this externalized research paradigm. However, is there sufficient standardization in data formats and metadata descriptions to support efficient information flow?

The Pistoia Alliance identified one specific example of an unmet need for such an information standard. There was no open standard for the computer-based definition of macromolecule structural information. As such the Pistoia Alliance launched its HELM project – the Hierarchical Editing Language for Macromolecules.

This talk will commence with a brief overview of the Pistoia Alliance as a member-driven, open innovation organisation, responding to the challenges of the information ecosystem. The Pistoia Alliance HELM project will be examined in more detail including the origins of this idea, an update on its current status, and a report provided on the increasing adoption of this standard – and the key players in the adoption story.

Ensuring the machine learner gets the right tuition and AI gets its nose; standardisation and benchmarking in diagnostic marker discovery in exhaled breath. Professor C.L. Paul Thomas, Centre for Analytical Science, Chemistry Department, Loughborough University.

There is a convergence of developments and capabilities that are driving the technology needed for non-invasive and rapid diagnosis and patient stratification. The potential of low-cost pervasive diagnostics is enticing with the prospect of earlier diagnosis, better management of care, and reduced care costs. However such a prospect remains firmly in the future. First molecular markers of injury, disease and distress must be characterised and translated into clinical practice, and then embodied into the “web of things”; integrated sensor systems that are reliable and affordable.

Examples of non-invasive breath based diagnosis are a good demonstration of this area of development as well as highlighting the complexity and difficulty of delivering reliable candidate molecular markers for trials, tests and development. Research in this space is driving measurements to the limits of currently achievable sensitivity with concomitant challenges for data processing and multi-variate modelling. The generation of such data demands a step change in the issues of proficiency, efficacy, standards, standardisation and benchmarking, right down to statistical process control within the breathing processes being sampled. These concepts will be studied through two case studies: one on sample storage and the other on benchmarking and proficiency.

Session 3: Information and Knowledge from Data

Building the utopian dream – toward the integrated, paperless drug discovery lab. Dr Allan Jordan, Head of Chemistry, Drug Discovery Unit, Cancer Research UK Manchester Institute.

Whilst the paperless laboratory has been forecast for several decades, progress toward its delivery has been slow.

Building a drug discovery team from scratch in 2009 has allowed us a rare opportunity to build an integrated informatics system from the ground up and has allowed us to think creatively about how we might finally deliver such an environment. However, this dream is not without its challenges. As a small team with limited informatics experience, integrating a centralised database with disparate items of lab equipment, on a limited budget, to deliver a useable system, has presented interesting challenges which we have sought creative ways to overcome.

This talk will describe the infrastructure we have implemented, or progress to date and some thoughts on where we would like to be in the near future. We will also discuss some of the problems encountered along the way and how we have chosen to address issues which are not uncommon to many drug discovery groups.

Using machine learning and cloud technologies for point-of-care in vitro diagnostics. Dr Omer Casher, Chief Data Officer, SIME Diagnostics, Stevenage Bioscience Catalyst.

The transfer of *in vitro* data analysis of bodily fluid samples from the R&D laboratory into a real world clinical setting is a complex challenge that can potentially be addressed using digital innovations not hitherto implemented in healthcare. In particular, machine learning algorithms, implemented in a public cloud infrastructure can be optimised for specific disease areas such as diabetes. The success of such a configuration is predicated upon effective data QC, automated workflows, and end-to-end regulatory compliance.

Enabling creative scientists to focus on science: automation, workflows, dashboards, insight, aha!!!. Martin Owen, Data to Insight Activist, GlaxoSmithKline.

For the past 15 years the climate that the pharmaceutical industry operates in has changed dramatically. Three-yearly cycles of reorganisation, relentless downsizing, and a more flexible workforce with increased outsourcing and temporary contracts comes at the cost of organisational learning, knowledge accumulation and knowledge sharing. At the same time, advances in automation and monitoring techniques push the bottleneck from experimentation to data processing.

The aim of this presentation is to show how the concept of “apps” can off-set this risk by providing more effective tools that are easier to use for both expert modellers and mainstream scientists and management. We make data more accessible, standardise workflows and interrogate Interactive dashboards to manage complexity, assess risk, evaluate control options and communicate insights.

The presentation will demo three apps in areas as diverse as biopharmaceuticals, formulation and accelerated stability in order to show the speed, simplicity and rigour that are possible using modern data analytics.

Session 4. How do we get There: Meeting Summary and Structured Discussion

Building the links. Mr Richard Kidd, Royal Society of Chemistry.

What other initiatives and activities are already running, and what new ones are needed? What role should the Royal Society of Chemistry have in connecting these together?

Measurement, Information and Innovation: Digital Disruption in the Chemical Sciences

Biographies

Speakers

Dr Danny Kingsley, Head of Scholarly Communication, Cambridge University.



Danny took up the position of Head of Scholarly Communication at the University of Cambridge in January 2015. She is overseeing all aspects of scholarly communication at the University, including compliance with funder open access policies, research data management, intellectual property, staff development, copyright and other areas. She also holds a Visiting Fellowship at the Australian National Centre for the Public Awareness of Science. Her research centres on scholarly communication with a specific interest in open access advocacy. She sits on the Australian Academy of Sciences National Committee for Data in Science. Danny's previous role was the Executive Officer of the Australian Open Access

Support Group which aims to inform the discussions around open access at a time of great change in this area. Her previous role was Manager, Scholarly Communication and ePublishing at the Australian National University where she was responsible for developing policies relating to scholarly communication and open access, and rebuilt the DSpace repository prior to its July 2011 relaunch. She also worked as an Associate Lecturer (part-time) in science communication. She completed her PhD at ANU in 2008 - looking at the barriers to opening up access to scholarly publications.

Dr Gerhard Noelken, Technology and Innovation Group, Pfizer Pharmaceutical Sciences.



Gerhard Noelken studied human medicine in Mainz (Germany) and worked as a scientist in toxicology, pharmacology and clinical research before becoming Head of High Throughput Screening for Aventis in Frankfurt. After a sabbatical in molecular biology with Biotech Australia he took over responsibility for Aventis R&D Informatics as a Frankfurt Site Head in 1996. Moving to Pfizer UK in 2001, Gerhard worked as Site Head for Discovery Research Informatics in Sandwich. Since 2007 he works in Pharmaceutical Sciences as the Global Business IT Lead. Gerhard's key interest areas are:

- Integration of informatics systems along the lab work processes in order to increase efficiency
- Implementation of a consistent data architecture as the foundation for collaboration and the scientific reuse of data (e.g. for prediction and simulation purposes).

Peter Boogaard, Industrial Lab Automation.



Peter Boogaard has extensive experience in laboratory management to enable cross-functional collaboration between research, development, quality assurance and manufacturing corporations. Peter is founder of Industrial Lab Automation and organizer of the Paperless Lab Academy congress. He is publishing in international magazines and contributes in several industry advisory boards. Peter is a Dutch citizen and studied Analytical Chemistry in Delft. www.industriallabautomation.com

Professor Krishna Persaud, School of Chemical Engineering and Analytical Science, University of Manchester.



Krishna Persaud, PhD, FRSC, FInstMC, graduated in Biochemistry at the University of Newcastle-upon-Tyne, UK and proceeded to a graduate with a MSc in Molecular Enzymology at the University of Warwick, UK and a PhD specialising in Olfactory Biochemistry. He subsequently worked at the University of Newcastle-upon-Tyne, University of Pisa and the Medical College of Virginia. He was appointed Lecturer in Instrumentation and Analytical Science at UMIST, Manchester, UK in 1988, and progressed to his current position of Professor of Chemoreception at the University of Manchester, School of Chemical Engineering and Analytical Science. He has been involved in research in chemoreception, crossing disciplines from biological aspects of olfaction to sensor arrays, electronics, signal processing and pattern recognition, and commercial development of artificial olfaction technologies. He has been involved in the development of gas sensor arrays for sensing odours based on conducting polymers, that became commercialised by Aromascan plc, and is currently a director of Multisensor Systems Ltd. He is Exchange Coordinator for the GOSPEL Network of Excellence, Committee member of the NOSE II forum, Executive Secretary and Treasurer of the European Chemoreception Organisation (ECRO), and is current President of the International Society for Olfaction and Chemical Sensing (ISOCS). He has been an associate editor for a number of journals that include *Materials Science and Engineering C*, *Biochemical Journal*, *IEEE Sensors Journal*, *Sensors and Actuators B*. He has received a number of awards that include the Prince of Wales Award for Technological Innovation Certificate 1995, European Space Agency Certificate of Achievement in 1997, Silver Medal Royal Academy of Engineering, 1999, Water Industry Impact Award 2015. He has over two hundred and twenty publications in the field of chemical sensing.

John Wise, Executive Director Business Development - Europe, Pistoia Alliance.



John Wise is an Executive Director of the Pistoia Alliance and the Programme Co-ordinator for the PRISME Forum. The Pistoia Alliance is a not-for-profit, cross-company organisation committed to lowering the barriers to innovation in life science R&D. The PRISME Forum is a not-for-profit pharma R&D IT leadership group focused on the identification and palliation of “hot topics”, and the sharing of industry best practices. John has worked in life science R&D informatics in a variety of organizations including academia, the pharmaceutical Industry, a cancer

research charity as well as in the technology supply side of the industry. This has provided him with direct, hands-on experience of delivering computer-based services across the life science R&D value chain. John graduated in physiology before obtaining a post-graduate certificate in education.

Professor C.L. Paul Thomas, Centre for Analytical Science, Chemistry Department, Loughborough University.



Paul joined Loughborough University in 2007 as a professor in Analytical Science to work within their Centre for Analytical Science team to develop new measurement approaches for non-invasive diagnostics and prognostics. Of particular interest has been the translation of ion mobility techniques (linear and differential) to in-vivo applications with an emphasis on emergency medicine.

As well as serving as Head of Department (2010 to 2015) he is also an Editor in Chief for the International Journal of Ion Mobility Spectrometry and a member of the international Advisory Board for the Journal of Breath Research. He is currently the Project Co-coordinator of the H2020 project TOXI-triage and Work stream leader (Ex-clinic analytics) in the MRC project EMBER.

Prior to joining Loughborough Paul gained his PhD in Analytical Science at UMIST and was awarded an endowed lectureship in the Department of Instrumentation and Analytical Science where he subsequently established a Volatile Organic Compounds research group. He remained at UMIST and subsequently the University of Manchester until 2007. During this time he was President of the International Society of Ion Mobility Spectrometry and an Associate Editor for the Analyst.

Dr Allan Jordan, Head of Chemistry, Drug Discovery Unit, Cancer Research UK Manchester Institute.



Allan Jordan joined the CRUK Manchester Institute Drug Discovery Centre in July 2009 as Head of Chemistry. He gained a BSc in Chemistry from UMIST in 1993 and, after a short spell as a teaching assistant in Arizona, he returned to UMIST to conduct post-graduate research into anticancer natural products. Following post-doctoral work at the University of Reading, he joined RiboTargets in Cambridge (now Vernalis) where he worked on a number of therapeutic areas at

various stages of the research pipeline. He has been involved in the delivery of seven pre-clinical candidates, of which three have successfully entered human clinical trials.

In his present role, alongside responsibilities for management of the chemistry team and progression of the drug discovery portfolio, he oversees the group's capabilities in structural biology, computational chemistry and chemoinformatics. With a strong focus on data integrity and accessible approaches to enhanced integration, the team were awarded an S-Lab Effective Lab award for IT and Informatics Integration in 2013.

He is also passionately involved in science communication and research engagement with the general public, fundraisers and technology providers through direct contact and the broadcast media. In 2014 he was a recipient of a Cancer Research UK "Flame of Hope" award (the charity's highest award for voluntary services) for his contributions to research engagement on behalf of the charity.

Dr Omer Casher, Chief Data Officer, SIME Diagnostics, Stevenage Bioscience Catalyst.



Dr Omer Casher is an entrepreneur, medical technologist and data scientist with 20+ years of related experience. He joined SIME Diagnostics with a mission to realise a goal of improving the early detection of disease indications using machine learning. His unique expertise in big data, clinical imaging informatics and regulatory compliance was developed at the GSK Clinical Imaging Centre (CIC) where he headed an informatics team supporting both early and late phase complex imaging clinical trials. After leaving GSK, he formed a company called IMAQA where for 3 years he channelled his skills to meet new big data challenges in real world evidence.

Martin Owen, Data to Insight Activist, GlaxoSmithKline.



Martin Owen graduated with a degree in chemistry at Liverpool University, joined Allen & Hanburys (now GlaxoSmithKline) and worked in the Medicinal Chemistry teams that delivered the Zantac and Imigran blockbusters. He then transferred to Chemistry Development twenty years ago where he pioneered the use of experimental design and automation and delivered modelling workshops at over 14 sites. He is currently an Informatics Incubation Leader in Product Quality Risk Management delivering transformational solutions to support quality by design across multiple scientific disciplines.

He is an active LeanSigma black belt practitioner and advocate of agile, lean start-up to application development. He is the GSK industrial supervisor for a statistics post-doc at Southampton University, an honorary Senior Lecturer at De Montfort University and has directed “Understanding Chemical Processes” workshops for academia for over ten years.

Mr Richard Kidd, Publisher, Royal Society of Chemistry.



Richard is responsible for the RSC’s developments in data publishing. He is Treasurer of the InChI Trust and is also involved in the IMI Open PHACTS project.

Session Chairs

Professor Jonathan Goodman, Department of Chemistry, University of Cambridge.



Jonathan Goodman studied boron-mediated aldol reactions during his PhD with Professor Ian Paterson FRS at the University of Cambridge. He then did a post-doc with Professor Clark Still at Columbia University, before returning to the chemistry department at Cambridge, where he is now Professor of Chemistry and Deputy Director of the Centre for Molecular Informatics. In 2013, he won the RSC's Bader Award. His research group does experimental chemistry, computational chemistry and data analysis.

Dr Helen Cooke, Chair, Chemical Information and Computer Applications Group, RSC.



Helen Cooke is Chair of the CICAG. A Chemistry graduate from the University of Manchester, she worked for 10 years, until 2014, at GlaxoSmithKline in the R&D IT department in Stevenage and Philadelphia, as an information manager and programme manager delivering IT solutions and improved processes for the benefit of R&D scientists. Previously, she was the MSc Cheminformatics course leader, lead trainer for the *CrossFire Beilstein* and *Gmelin* databases, and the Chemistry Librarian at the University of Manchester Institute of Science and Technology. In her early career she worked for the International Union of Crystallography and Derwent Information. Helen has a strong interest in the history of chemical information, which was the subject of her PhD.

Mr John Trigg, Automation and Analytical Management Group, RSC.



John Trigg is the Vice-Chairman of the Automation and Analytical Management Group of the Royal Society of Chemistry, member of the Environmental, Sustainability and Energy Division of the Royal Society of Chemistry, and an Advisory Board Member of the Institute for Laboratory Automation. He has extensive experience in the field of R&D data, information and knowledge management, he has been author of a number of publications and booklets on electronic laboratory notebooks and laboratory informatics, and has presented papers and run workshops at conferences in the UK, Ireland, mainland Europe, USA and Australia. He was the recipient of the 2000 International LIMS Award.

Professor Jeremy G. Frey, Department of Chemistry, University of Southampton



Jeremy Frey obtained his DPhil on experimental and theoretical aspects of van der Waals complexes at Oxford University, followed by a NATO fellowship at the Lawrence Berkeley Laboratory. In 1984 he took up a lectureship at the University of Southampton, where he is now Professor of Physical Chemistry. His experimental research probes molecular organization in environments from single molecules to liquid interfaces using laser spectroscopy from the IR to soft X-rays. He investigates how e-Science infrastructure can support scientific research with an emphasis on the way digital infrastructure can support the intelligent access to and analysis of scientific data.