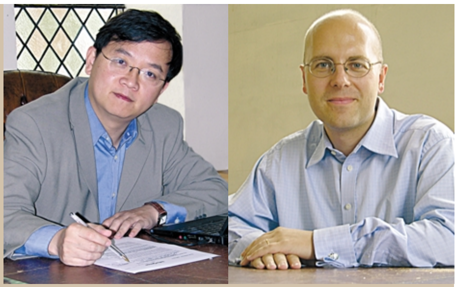
interview

Yike Guo and Jonathan Sheldon of InforSense discuss the impact of workflow technology on drug discovery

Interviewed by Christopher Watson



Yike Guo,
CEO & Founder, InforSense
Jonathan Sheldon,
Chief Scientific Officer,
Life Sciences and Healthcare

Yike Guo founded InforSense in 1999 to commercialize his group's pioneering open workflow technology for high-performance large-scale integrative data analysis, rapid application building and process knowledge management. He has led the company's growth since then. He is a world leading expert in large-scale data mining and grid computing and also serves as Technical Director of the Parallel Computing Center and Head of the Data Mining Group at Imperial College, University of London. Over the past four years he has led a number of significant academic and industrial research and development projects targeted at building next generation e-Science platforms for which he has gained UK and European funding in excess of £10 million. He holds a PhD in computing science from Imperial College.

Jonathan Sheldon is responsible for managing InforSense product development and delivery to meet customer needs in life sciences and healthcare. Prior to InforSense he was Chief Technology Officer for Confirmant, where he was responsible for developing the company's proteomics products and services. Previously he established the first bioinformatics group and was Head of Bioinformatics for five years at Roche Welwyn, UK, participating in a number of global initiatives within the company. Sheldon holds a PhD in molecular biology and biochemistry from the University of Cambridge.

Could you give me some information about InforSense?

YG: The Company was founded in 1999 and was spun out from Imperial College, London. Our business is centred on developing and marketing our integrative analytics platforms and our current business focus is the life science industry. Our approach uses a workflow paradigm as a generic mechanism to integrate different databases, software applications, web services and shared expertise to improve decision-making across an organisation. The important thing is that we have developed an environment that allows life science researchers to integrate data resources and applications themselves without involving any programming. This forms the bedrock of what we do.

For the benefit of readers who are unfamiliar with the concept, could you explain what a workflow is?

JS: Workflows are not really a new concept and have been widely used in business process management for a while, but they are now becoming an important technology for life science discovery informatics. For us a workflow is a mechanism to integrate data, applications and services. Workflow technology enables scientists to dynamically construct their own research protocols for scientific analytics and decision-making by connecting various information resources and software applications together in an intuitive manner. Our customers use workflows in pretty much all the major approaches within drug discovery: gene expression, sequence analysis, proteomics, library design and so on. The technology is generic, so analytic workflows can be built for any area. There are also appropriate interfaces so that interactions can be formed between all the major scientific software providers. Therefore, software from a variety of vendors can be assembled into something that is the ideal workflow for the end-user.

YG: This is a key point. This approach is able to deal with all the different data types, algorithms and data sources across the whole drug discovery process. The platform is generic and layered onto this platform is the ability to work

interview

within and across particular domains such as bioinformatics, cheminformatics and also clinical and biomedical informatics. Many areas in the discovery process have become compartmentalized and because a platform such as this is so generic it enables biochemists, chemists and others involved in drug discovery to sit around a table and look at the data at an appropriate level so that they are all comfortable with what they know to make decisions.

Can you talk about some of the pharma companies that have implemented these workflow technologies?

YG: In the last year more major pharma companies and leading research institutions have deployed our technology. This has been across a variety of different fields – in chemistry, biology and clinical informatics – and these companies are now appreciating the commonality and functionality of this integrative analytics approach. For example, GlaxoSmithKline (GSK) is using InforSense technology to provide a workflow-based decision support environment for discovery scientists. Their researchers are accessing existing in-house cheminformatics tools, combining data sources, methods, tools and web services, and integrating them as needed. The InforSense Portal provides rapid development and deployment of web-services, generated automatically from InforSense workflows, to a large community of scientific users. The two companies also recently announced a three-way collaboration that also involves Spotfire to move this collaboration to the next level, enabling GSK to manage, orchestrate, and deliver to end-users applications at the pace of discovery.

JS: It is also worth saying that because this technology represents process knowledge you can look at workflows at various levels of detail. One can analyse how many compounds there are in the various stages of drug discovery and drill down to an individual workflow, for example, which suggested that this particular compound in this particular study was the one to take forward into the clinic. It is a way of working with all the different levels of drug discovery; encapsulated in the concept of 'bench to boardroom'.

What are the key benefits that workflow technologies bring to drug discovery informatics?

YG: There is traditionally a large gap between information technology (IT) people and scientists – IT can be perceived as a bottleneck in scientific research. Whenever one needs to do any integration, IT people may need to be consulted or software has to be bought. Workflows offer scientists the ability to access a variety of software packages and data sources and to construct their own integration processes without having to manually transform data or program interfaces; essentially using the technology as infrastructure. This is our fundamental value proposition.

Could you briefly discuss some of the collaborations you have?

YG: Partnerships are vital to InforSense. Our customers need to access their choice of data and specific domain-focused tools via our integrative analytics to solve their problems. So, for InforSense the strategy is to partner with best-of-breed software and infrastructure providers, ensuring that their software and solutions are compatible so that customers can use InforSense to integrate the data and tools they need. For example, we have been working closely with Oracle for some time to provide in-Oracle data processing and analytics, offering advantages for data security and performance. We already mentioned our partnership with Spotfire and we also have proven integrations with other leading vendors and consultancies.

Information analysis and meaningful integration of data is particularly relevant to systems biology. How are InforSense positioned to address the challenges of systems biology?

JS: The very foundation of systems biology is to bring together heterogeneous data types, including sequence, genomic, proteomic and metabolomic information. Combining large volumes of data, a variety of data formats and appropriate tools across domains is still a challenge and is the first step in building accurate cross-discipline models, which can easily be deployed across an organization. There is clearly a broad data integration element to this, which workflow approaches can accomplish. The Windber Research Insitute (WRI) is deploying InforSense technology to provide a workflow-based informatics platform for translational medicine. The WRI has many high-throughput experimental techniques that generate a lot of data and they also have access to a large amount of detailed clinical data, which is unusual. This is mainly through the strong ties they have with medical centres in the US, particularly in their immediate location. This huge resource of clinical data has enabled them to adopt a very top-down, physiology driven approach to systems biology. This involves integrating this clinical data with high-throughput experimental data and then attempting to build models from this data, such as predictive models of breast cancer disease progression. Such a model is the result of many workflows, which can be deployed out over the web using portal technology to physicians, primarily in this case, who can then access these workflows. Therefore, via a few clicks of the mouse one has instantly gone from what may be a complicated workflow to a simple web application where a physician may upload a test, run a workflow in the background and then a therapy recommendation can be deployed out through the web again. It is very much a case of going from bench to bedside but, because the clinical data play such a central role, it also goes the other way, from bedside to bench.

Where do you see yourself and the company in five years and then ten years from now?

YG: Our technology is rapidly being adopted in the market place. In the next two to three years we aim to consolidate our position in the life science field and to provide a *de facto* standard platform for discovery informatics. In certain fields, such as systems biology and translational medicine, this flexible integration power represents a cutting edge approach. Our goal is to cover the whole value chain of drug discovery, in essence to provide a discovery operating system.

Yike Guo *and* ***Jonathan Sheldon***
InforSense Ltd,
459a Fulham Road,
Chelsea,
London SW10 9UZ,
United Kingdom