Institution: Imperial College London

Unit of Assessment: 11 Computer Science and Informatics

Title of case study: Case Study 5: Knowledge Management Technology for Pharmaceutical and Healthcare Industries (InforSense)

1. Summary of the impact (indicative maximum 100 words)

The research in this case study has pioneered knowledge management technology. It has had major impact on drug discovery and translational medicine and is widely adopted in the pharmaceutical and healthcare industries. The impacts are:

1. The formation of InforSense to commercialise the technology. The company had 150 employees in June 2009 when it merged with IDBS Ltd to create the world's second largest life science informatics company.
2. The results from knowledge management technology and associated software platform have enabled the integration of molecular, imaging, clinical data and analytics, to identify biomarkers for disease identification, treatment selection and side effect prediction.
3. Since 2002 the technology has been deployed by major pharmaceutical companies (including GSK, AZ, Roche, Pfizer, Bayer and Boehringer Ingelheim) and leading healthcare institutions e.g. Mayo Clinic, Harvard Medical School and King's Health Partners, generating significant social, health and economic impact.

2. Underpinning research (indicative maximum 500 words)

The underpinning research has been carried out in the Discovery Science Group, Department of Computing, Imperial College London. The group is led by Professor Yike Guo who joined the faculty in 1997.

The research of the group has been focused on advanced software technology for large-scale data analysis. In 2001, the group started the Discovery Net UK e-Science Pilot Project funded by a major EPSRC grant [i]. In this 3.5-year project, the group developed the world's first Grid-based Collaborative Knowledge Discovery and Management Platform [1]. This facilitated integration of a set of component systems or services to form a workflow to enable scientists to create data analysis applications using multiple sources of data. The core of the developed technology is a novel scientific workflow model allowing end-user scientists, not programmers, to dynamically search and visually construct data sources, manipulate and provide analysis services, and then compose them into workflows. This technology proposed and implemented in 2002, was the first analytical workflow technology for large scale distributed data analysis in the world and won the "*Most Innovative Data Intensive Application Award*" at the major international conference in this area, Supercomputing 2002.

Since then we have continued to develop the technology into a Cloud computing model for supporting global scientific collaboration for large-scale data analysis [ii]. The innovation includes: dynamic information structuring, allowing users to access and integrate the required heterogeneous data sets on the fly in the workflows by transforming generic queries into the language that is specific to a particular data set [4]; scalable knowledge discovery and data mining tools for terabyte scale data (big data) [2]; and intellectual property management in collaboration via tracking the provenance of knowledge discovery processes in scientific research [iii]. Thus, the technology provides a complete process of e-science by analysing distributed and heterogeneous data sets for global scale research collaboration. This research pioneered many key areas such as service oriented workflow [3], dynamic service generation and deployment [2], big data analysis [iv], open science infrastructure [iv, vi], all of which are now becoming mainstream computing. In the subsequent Discovery Science platform grant [ii], the technology has been integrated and further extended into the development of the IC Cloud [5] with the first Big Data architecture developed for many important open research activities such as smart cities and translational medicine [iv]. With these applications, innovative research such as elastic algorithms [v], collaborative sensing [iv] and big data technology [iv, v] for translational informatics [vi] are proposed and pursued, resulting in significant academic and economic impacts.

3. References to the research (indicative maximum of six references)

Publications that directly describe the underpinning research

* References that best indicate the quality of the research

[1] *A. Rowe, D. Kalaitzopoulos, M. Osmond, M. Ghanem and Y. Guo. The Discovery Net system for high throughput bioinformatics. Bioinformatics, 19(Supplement 1): 225-231, 2003. (http://dx.doi.org/10.1093/bioinformatics/btg1031)

[2] M. Ghanem, V. Curcin, P. Wendel and Y. Guo. Building and Using Analytical Workflows in Discovery Net. In Data Mining Techniques in Grid Computing Environments (editor W. Dubitzky), John Wiley & Sons, 2008. (http://dx.doi.org/10.1002/9780470699904.ch8)

[3] *S. AlSairafi, F. S. Emmanouil, M. Ghanem, N. Giannadakis, Y. Guo, D. Kalaitzopoulos, M. Osmond A. Rowe, J. Syed and P. Wendel. The design of Discovery Net: towards open grid services for knowledge discovery. The International Journal on High Performance Computing Applications, Special issue on Grid Computing: Infrastructure and Applications, 17(3): 297-315, 2003. (http://dx.doi.org/10.1177/1094342003173003)

[4] *N. Giannadakis, A. Rowe, M. Ghanem and Y. Guo. InfoGrid: providing information integration for knowledge discovery. Information Sciences, 155(3): 199-226, 2003. (http://dx.doi.org/10.1016/S0020-0255(03)00170-1)

[5] R. Han, L. Guo, M. Ghanem, M. Osmond and Y. Guo. Enabling Cost-Aware and Adaptive Elasticity of Multi-tier Cloud Applications. Special Issue on Cloud Computing, Journal of Future Generation Computer Systems, 2012. (http://dx.doi.org/10.1016/j.future.2012.05.018)

Grants that directly funded the underpinning research

[i] Discovery Net: An e-Science Test Bed for High Throughput Informatics. EPSRC GR/R67750/01. Y. Guo (PI), £2,082,704, October 2001 – March 2005.

[ii] PLATFORM: Discovery Sciences Research Group: Applying Real-time Data Mining for Large Scale Scientific Applications. EPSRC EP/C53492/1. Y. Guo (PI), £ 409,411, October 2005 – September 2010.

[iii] U-BIOPRED: Translational Medicine for Airway System Disease. EU-IMI. Y. Guo (CI). €20,685,241, June 2010 – May 2014.

[iv] Digital City Exchange. EPSRC EP/I038837/1. Y. Guo (CI), £5,930,480, October 2011 – September 2016.

[v] Elastic Sensor Networks: Towards Attention-Based Information Management in Large-Scale Sensor Networks. EPSRC EP/H042512/1. Y. Guo (PI), £471,777, June 2010 – December 2013.

[vi] eTRIKS : European Translational Informatics and Knowledge Management Services EU-IMI. Y. Guo (PI). € 23,700,000, October 2012 – September 2017.

4. Details of the impact (indicative maximum 750 words)

Economic impact

InforSense was formed as a spinout company from the Department of Computing, Imperial College

London in 1999 by Professor Yike Guo. In 2002, Imperial College assigned the intellectual property rights of the technology developed in the Discovery Net e-Science Pilot Project to InforSense for commercialisation. The company was merged with IDBS in June 2009.

InforSense: In 2003, after winning the "*Most Innovative Data Intensive Application Award*" at Supercomputing 2002, the company put in place its first organized sales force, with a focus on Life Sciences bringing it the first pharmaceutical customer (GSK). In 2007, InforSense was ranked amongst the top 25 fastest growing private technology companies in the UK by The Sunday Times and was included in the 2008 Red Herring Finalist of Top 100 Companies in Europe. The company grew from 5 to 150 employees between 2002 and 2009, with a customer base of nearly 100, 70% of which are Fortune 200 companies and all major pharmaceutical companies (including GSK, AZ, Novartis, Roche, Bayer, Pfizer, J&J, Ely Lilly). InforSense generated over £15M sales before merging with IDBS in June 2009. IDBS became the world's second largest life science informatics company, and then became the world-leading provider of translational informatics solutions. [A]

IDBS: The IDBS healthcare informatics technology is directly based on InforSense's technology. The merger allowed IDBS to start its healthcare informatics business with 8 large healthcare organizations as new customers [A, B, H]. The 2010 Company Report of IDBS [G] shows a 29% increase in revenue since the takeover. The company now has 275 employees worldwide and revenues of $50M with subsidiaries in UK, USA, France, Japan and China. Prof. Guo has been the Chief Innovation Officer since the merger. Through IDBS [C], the technologies such as analytical workflow, research provenance management and collaborative support, developed in the underpinning research within the Department are currently used by more than 200 pharmaceutical companies, major healthcare providers, global leaders in medical research, and high tech companies to manage and analyse large scale research data for industrial R&D and clinical research [A]. IDBS won the Queen's Award for International Trade 2011 for outstanding business performance and technology innovation.

Pharmaceutical and Healthcare Industry Impact

The main impact of the developed technology is felt in the area of drug discovery and translational medicine research:

1. Drug-discovery research where genomic, proteomic and metabolomic data have to be integrated with chemical information, imaging and textual data in the same analysis pipelines, with the aim of discovering and developing new drugs [1], [F].
2. Translational medicine research where the analysis of the genomic, proteomic and metabolomic data needs to be integrated with patient data and medical records with the aim of identifying disease biomarkers, selection and design of treatment protocols and prediction of side effects for healthcare and for future personal medicine [D, F].

Within the pharmaceutical industry, the knowledge management technology developed in the department has been used in most major pharmaceutical companies through various InforSense and IDBS products. Within the domain of healthcare applications, the technology is currently used at the ***Dana Farber Cancer Institute of Harvard Medical School*** to integrate and analyse patient and sample data to define cohort studies in cancer genomics studies [D], and at the ***CHOP (Children Hospital of Philadelphia)*** to support Genome Wide Association data analysis integrating patient data and genotyping results. It is used at the ***Mayo Clinic*** to support the development of personal medicine. ***Erasmus Medical Centre*** and ***Southampton University Hospital*** have used it to develop new treatment methods for Acute Myeloid Leukemia and Asthma. ***Windber Institute/Walter Read Army Research Centre*** used the technology to build the first completed translational breast cancer research database, covering all women in the US Army. The system has been used to study life style, cancer prevention and determine effective treatment for all female soldiers in the US Army. It was the largest translational research project in the US army in 2010 [D]. It was also used at ***King's Health Partners*** as the basis of its Oncology Research Information System (ORIS) for large-scale translational research. The initial ORIS deployment enables combining clinical, genetic and tissue sample data across more than 26,000 historic breast cancer patients alongside a current feed of new, consented patients' data direct from the clinic into the longitudinal research database [E].

The ***National Centre for Mental Health (NCMH)*** in Wales established the Wales Mental Health Network (WMHN) to recruit 6,000 volunteers for studying mental health disorders such as schizophrenia and bipolar disorder. The workflow technology and the big data collaborative framework based on the underpinning research enable clinical cohorts to be easily defined and analysed from new subjects as well as providing access to over 3,000 historic records.

The success of the application of the technology led to major EU funding as part of the €24M eTRIKS project from IMI in Oct. 2012 a 5 years project to build up a translational informatics cloud for Public Private Partnership-based clinical trials. The underpinning technology is the core component of this industrially led project for which Prof. Guo is the academic PI. The project involves the participation of 12 major pharmaceutical companies and medical research institutions – Roche, AstraZeneca, Sanofi, Pfizer, Merck, Lundbeck, Janssen, GSK, Lilly and Bayer have provided €11.5M funding and committed to make eTRIKS the industrial common platform for translational research. This activity is stimulating the formation of the tranSMART foundation as a global organisation to standardize translation informatics technology. tranSMART's technology originated from InforSense's translational informatics technology. Prof. Guo has been appointed as the Chief Technical Officer of tranSMART foundation [F]. The first version of the full open source tranSMART system (tranSMART – eTRIKS version) was released in June 2013 and deployed in 6 major pharmaceutical companies and many research institutes worldwide.

5. Sources to corroborate the impact (indicative maximum of 10 references.)

[A] Chairman and CEO of IDBS confirming details regarding InforSense and IDBS

[B] IDBS Press Release on the acquisition of InforSense by IDBS
http://www.idbs.com/news/PR/09jun26.asp Archived on 18/11/2013
https://www.imperial.ac.uk/ref/webarchive/07f

[C] IDBS Press Release describing the use of InforSense technology in its products
http://www.idbs.com/Data-Management-News/press-release/10FEB23.asp
Archived on 21/10/2013 https://www.imperial.ac.uk/ref/webarchive/zyf

[D] S. A. Beaulah et al. Addressing informatics challenges in Translational Research with workflow technology. In Drug Discovery Today. 13(17–18): 771–777, 2008.
http://dx.doi.org/10.1016/j.drudis.2008.06.005

[E] Kings Health information
http://www.laboratorynetwork.com/doc/IDBS-Delivers-Platform-To-Accelerate-0001
Archived on 21/10/2013 https://www.imperial.ac.uk/ref/webarchive/1yf

[F] tranSMART foundation information
http://www.transmartfoundation.org/site/about-us/our-management-team
Archived on 21/10/2013 https://www.imperial.ac.uk/ref/webarchive/2yf

[G] IDBS 2010 Company Report - available on request

[H] Press Report on IDBS Success in integrating InforSense technology available at
http://www.genomeweb.com/informatics/idbs-releases-mid-year-results-says-it-track-best-ever-financial-performance-201 or available on request