

Privacy-preserving analytics will go mainstream

By Bruce McCabe

First published in *The Australian* 9 August 2005

Copyright S2 Intelligence Pty Ltd 2005



In future, business intelligence systems will chat to their friends and neighbours as they go about their work.

What I mean by this is that one day Australian corporations can expect to find themselves using the business analytics software provided by companies like Hyperion, Cognos and SAS to analyse not only the data within their four walls, but also vast amounts of valuable and relevant data from business partners. Indeed, they will quickly reach a stage where the external data they analyse will be far greater than the data they own.

Web services are one of the catalysts for this, because they make inter-organisational connectivity vastly easier and cheaper to achieve. Semantic Web developments—aimed at adding structure and metadata to web content to improve machine-navigability—are another.

New innovations developed by Australian researchers are also set to make a vital contribution.

First, a little background. There can be no doubt about the value of data sharing between trading partners. Manufacturers depend on information from wholesalers and retailers of their products, banks rely on information from mortgage brokers and property developers they work with, and car companies rely on information from dealerships. And there is, of course, a whole industry built around web-merchants that buy, sell and rent out their accumulated data.

The last example in particular should instantly remind us of the privacy issues that apply to data sharing. These represent a significant and ever-present hurdle for ethical corporations that wish to share information with business partners while respecting the privacy of their customers and staying within the guidelines set by government.

Securing privacy is critical, but acts as a brake on data sharing wherever customer transactions and information are involved in any way. Data must be carefully cleaned, sanitised or reprocessed by a provider before being passed to the recipient. Such delays make real-time analysis of shared data a pipe-dream.

It is this challenge that scientists at Australia's CSIRO are helping us overcome.

In June last year I wrote about two scientists, Dr Christine O'Keefe and Dr Kerry Taylor, and their involvement in a project to connect disparate sources of health-related data located in hospitals, universities, government departments and pharmaceutical companies. This initiative, entitled the "Health Data Integration project" set out to eliminate bottlenecks faced by medical researchers as they go about trying to correlate data from many institutions in their efforts to develop new treatments, medicines or preventative measures. HDI is achieving that objective by building ready to go electronic links and—here is the important part—providing electronic mechanisms so each and every data custodian can control which parts of their data can be analysed and what analysis can be performed.

As at mid 2005, the HDI project has been applied to lung cancer research in Queensland and is being picked up by a number of other healthcare communities of interest. One day I expect it will underpin something resembling a national health data network in this country.

But its usefulness goes beyond healthcare: this innovation will facilitate real time data sharing in every industry. If the technology is robust enough for the ultra-privacy sensitive world of patient and health data then it will be more than secure enough for any other context.

Two weeks ago I had the opportunity to address an audience of business people at the Cognos Australia's annual customer forum, where I discussed HDI as one of a number of leading-edge projects that illustrate what the future holds for mainstream business analytics.

In one of those wonderful bits of timing, the following day I received news from CSIRO that it plans to extend its HDI innovations to analytics in sectors such as banking and finance. They have now coined the term "Privacy-Preserving Analytics" (PPA) to describe the technology. I am not fond of new tech-acronyms, but it has to be said that this is a pretty catchy one.

Dr O'Keefe also confirmed that her organisation is now pursuing commercialisation. And well it should: the CSIRO team has developed something that is not only delivering real value in healthcare today, but could deliver substantial value in hundreds of scenarios.

Someday soon, either PPA or a derivative of it will be imbedded in mainstream analytic software products to help business partners in every sector analyse shared data in a more efficient, secure and timely way.

Bruce McCabe is an independent technology analyst and managing director of S2 Intelligence.

mccabe@s2intelligence.com.au