

DATA POLLUTION

Information integration may seem a worthy pursuit, but it has a dark side that Graham Oakes has only just seen.



Graham Oakes: as we integrate data, we make it more vulnerable

Technology guru Bruce Schneier once said: “Data is the pollution of the information age. It’s the by-product of every computer-mediated interaction. It stays around for ever unless it’s disposed of. Just as we look back at the beginning of the previous century and shake our heads at how the titans of the industrial age could ignore the pollution they caused, future generations will look back at us.”

I’ve spent much of my life trying to integrate data. Customer data in one system needs to connect to product data in another, so we can deliver a decent service. Plans in one system need to connect to specifications in another so we can build the right product.

Or, as often as not, customer data in a dozen different systems needs to be linked together so we can simply count how many customers we have.

Such integration has always seemed valuable. By enabling staff to see the big picture, we can help them to make better decisions, to work together more effectively, to eliminate duplication, and so on. It’s not always easy, but it has always seemed valuable.

Now it’s becoming clearer that there are larger costs to data integration – beyond the costs of data cleansing and transformation, of integration hubs and buses.

Data integration creates serious concerns for people’s privacy, both by increasing the amount of personal data which is accessible in any one place, and by increasing the ease with which people can access this information.

What’s valuable for the company which holds the data may be threatening to the people that data describes.

Data integration also dramatically increases the impact of data losses. At a time when each week brings fresh news of such losses – disks go missing in the post; laptops get left on trains; credit card processors get hacked – this is a serious concern.

We can tighten up our information-handling processes. We can train our people better. But mistakes will always happen. No system is infallible. Integrated data just increases the cost of these failures.

As we integrate data, we don’t just make it more valuable: we also make it more vulnerable.

I don’t think we can ignore these costs any longer. It’s no longer enough to seek ways to integrate data across multiple systems.

Even as we do this, we need to design in ways to partition that data so that it can’t be connected except by authorised people and for purposes for which we have demonstrably obtained consent.

Data pollution is like the worst type of chemical smog: each leak doesn’t just add to the smog, it interacts with what is already out there and multiplies the effects. It’s time we started to manage it seriously.

● SOA, Web Services & Enterprise Integration Evaluation Centre Expert Dr Graham Oakes is the principal of content management, product development and customer service strategies consultancy Graham Oakes Ltd. Email: graham@grahamoakes.co.uk. Website: www.grahamoakes.co.uk.