

Information Overload

Issue 67, May 2008

Welcome:

Previous editions of Information overload has touched on the subject of our digital selves, that is, information that is stored on a variety of electronic devices as well as information that we generate as a result of being “online”. Today I would like to take that a step further and discuss the problems we face as record keepers with our expanding digital shadows and the impact that has on privacy.

Lorraine Bradshaw
Marketing Coordinator and Projects Officer

In this Issue we will be looking at:

- The storage issue and its impact on record keeping requirements
- Storage and privacy issues
- The numbers
- A Thought to ponder.

The storage issue: and its impact on record keeping requirements:

According to the research conducted by IDC, 2007 was an auspicious year. It was the year where we created more information digitally than we could actually store. *The diverse and exploding digital universe. An updated forecast of worldwide information growth through 2011. An IDC White paper, March 2008.*

Granted there is a case that says we don't actually want or need to store every single entity created in the digital world, but it is a major problem nonetheless – namely what **do** you keep vs. what **should** you be keeping? Given that emails, text messages and other seemingly transient data forms such as search histories have been requested in courts of law it is a little worrying. And yes I did say search histories - you know those idle moments you spend surfing the web can and have been subpoenaed by organisations such as the US Government in its bid to find out just what you've been up to., *Hiawatha Bray, “Google Faces Order to Give Up Records,” The Boston Globe, March 15, 2006.*

Storage and Privacy issues:

It has been reported recently that a simple click on an embedded link within a website and / or email can have the FBI knocking on your door, removing your computer and throwing

© IEA 2008. All rights reserved. You are free to use material from the Information Overload newsletter in whole or in part, as long as you include acknowledgement of source.

Information Enterprises Australia Pty Ltd
Unit 4, Upper Level, 201 High Street, FREMANTLE WA 6160
Tel: 08 9335 2533 Fax: 08 9335 2544 e-mail: training@iea.com.au

you in jail. “But I didn’t click the link your honor” is apparently not much of a defence, but unless your office has CCTV footage of the exact moment of the infringement – something tells me you’d better get a good lawyer!

http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9080379&source=NLT_PM&nid=8 (accessed 08.05.08)

Having just re-read that last paragraph, I casually mentioned CCTV footage in our offices – no IEA doesn’t have CCTV – but other organisations do, and sometimes they don’t know about it. For example, a recent news story focussed on a camera that had been embedded in the ceiling of the female changing rooms at SBS offices in Sydney - <http://www.abc.net.au/news/stories/2008/05/06/2236386.htm> with the subsequent photographs being found on a computer being used by an employee.

Whilst the SBS camera may be an isolated incident as far as illegal camera placement is concerned, there are huge numbers of legitimate cameras that can and do capture us as we go about our day-to-day activities. You can and will be photographed in disparate places such as your corner shop to airports, crossings, ATM’s, petrol stations to taxis. This CCTV footage forms a considerable part of our digital shadow. That is, digital information that is captured about us, but not generated by ourselves. And can be captured

But our digital footprint is not just limited to images. Consider things like credit-card trails of purchasing behaviour, (how do you think Amazon knows what to recommend to you?) as well as names on mailing lists, and as we have mentioned our internet searching history, even this newsletter will form part of our shadows. Every key stroke that is captured, overwritten, changed, saved and sent out – can be followed and evidence of behaviours determined. This is both interesting and a little scary.

Of course our digital shadow also contains things that we do during a normal day. For example, the simple act of opening a web page, creating and downloading a file (document, video clip, picture file etc etc) all leave a digital footprint that can be followed, should someone be keen to know what you’ve been up to during the day. A case in point - I remember one place I worked; my colleague (who shall remain nameless) had the job of creating the monthly bulletin. The person always named the document something strange as it was being created, changing the name of the file before sending out....until the day it wasn’t changed of course and we received a phone call from the then CEO. Thankfully he was amused....

Looking at our online searching habits in a little more detail, we can see where the big marketing pushes are going to come from. Do you have a Google Mail (gmail) account? Google mail adds “ads” to each message. One of the interesting things I have noticed as I have opened emails from various entities across the world is how relevant the ads are that are served up alongside them. By relevant I mean, it picks up on the organisation, the subject line as well as the body of the message to deliver ads it thinks you might be interested in. OK it gets it wrong now and again, but that’s what makes it interesting. Namely the way the algorithms are set to pick up on the data you give it, and then translate that into potential revenue for someone else. Believe me when I say people are making a tidy living out of buying ad word ads.

When it comes to privacy issues, if we don’t know what is being captured about us, how can we ensure our privacy? Should we be more concerned about our privacy? Or are we “happy” to lose some of our anonymity to gain value added services, such as advertisements to services and products that we may be interested in?

As a research librarian and a web developer I am interested in both functionality and security of data. What is interesting though is the number of people who create sites that allow you to gain access to privileged information about organisations personal information. Last week I uncovered the pay scales for teachers (including names and positions) for a university in Americas mid west. The Problem is – once this kind of information has reached the realms of the search engines it cannot be stuffed back behind firewalls – sssh no-one was looking. And the reason is the digital footprint. Each website is cached into memory, and those pages that are easily found and therefore read by the automated arachnids are indexed and stored on numerous servers. And as I found this particular document it also forms part of my digital footprint as well.

Concerned? Yes it can be a huge worry. When you are giving your personal information to a third party, you can only hope they know how to store and secure your information properly. A report yesterday by Computerworld indicated that over half a million websites running phpBB open source software have been hacked - http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9084991&intsrc=hm_list

Unfortunately organisations aren't learning from other people's mistakes. After the grocery chain TJX was hacked and thousands of credit cards (and PIN's) were stolen, Hannafords fall to the same kind of attack comprising over 4 million card details. What was interesting was that – the credit card details were stolen as the data was transferred to the banking institutions...in an unencrypted file format. And they're the ones we hear about.

Apart from buying everything cash, running a stand alone computer with no internet access at all, is there anything we can do?

The Numbers:

Unfortunately not, if the data is anything to go by! We are using the net for more things than ever before. A study conducted by comScore in December 2007 noted that Yahoo had the potential to gather data through 400 billion events during the month. Time Warner, which includes AOL, was second, with about 100 billion events. Google was not too far behind with 91 billion... <http://bits.blogs.nytimes.com/2008/03/09/how-do-they-track-you-let-us-count-the-ways/> this study was conducted in just the United States. Imagine how big the digital shadow would be had they tracked the entire online world population. For your information - these "events" include - pages displayed, search queries entered, videos played, and advertising displayed within a page.

400 billion events in one month alone! Think about the record keeping system needed to manage that kind of traffic, it's quite staggering when you think about it. Then it shouldn't come as any shock that "in April 2007, for the first time, the total amount of information generated worldwide—some 281 exabytes (an exabyte is 1,024 petabytes, and a petabyte is 1,024 terabytes)—exceeded the capacity of all the hard drives, tapes, CDs, DVDs, and volatile and non-volatile memory created to hold it, according to IDC." <http://www.xconomy.com/2008/03/11/idc-tracks-your-digital-shadow/>

The main findings of the "Digital Universe" study:

- The digital universe is bigger than we thought: 281 exabytes in 2007 — 10% bigger than we said it would be last year. Digital cameras and security cameras out-shipped our expectations.
- By 2011, there will be 1,800 exabytes, or 1.8 zettabytes, of electronic data in existence. (NB given the last estimate was out by a considerable amount — something tells me this will be a gross understatement of the digital information created every year, especially as new programs requiring more processing power are generated every 5 minutes (or so it seems)).
- The digital universe will grow tenfold in five years.
- While 70% of the information in the digital universe is created by individuals, enterprises have some responsibility or liability for 85% of it (think of Google getting sued by [Viacom](#) for \$1 billion because of the videos uploaded by consumers).
- There is now more information created in a year than there is capacity to store it.
- Responsibility for the information in the digital universe by industry does NOT match IT spend or GDP distribution by industry.
- A single e-mail with a 1MB attachment sent to four colleagues can generate 50MB of information in the digital universe (based on an IDC-like architecture).

Taken from: "The diverse and exploding digital universe. An updated forecast of worldwide information growth through 2011. An IDC White paper", March 2008.

With many thoughts
Lorraine

A Thought (or two) to Ponder:

"There is no reason for any individual to have a computer in his home"

Ken Olsen, Digital Equipment, 1977

&

"I think computer viruses should count as life. I think it says something about human nature that the only form of life we have created so far is purely destructive. We've created life in our own image."

Stephen Hawking