



"turning data into dollars"

Tom's Ten Data Tips – July 2006

Privacy

Privacy is one of those topics that nobody cares about until their own privacy is being violated. Privacy threats have been compared to George Orwell's "1984" where a totalitarian regime decimated individual freedom. *Nowadays*, the privacy threat doesn't come from communist states but from capitalism, free markets, exchange of digital information and smart use of advanced technology.

In the 'off-line' world we've learned that sustainable industrial development doesn't compromise the environment, it must preserve it. Or else there wouldn't be enough employees and customers around to maintain a market standing. By the same token, the conveniences of our increasingly electronic world should not threaten privacy. Instead, control over disclosure of personal information is a necessary prerequisite, or progress in our digital economy will come to a standstill.

1. Technology forms *no* threat, *people* do

Technology is often discussed as if it were an autonomous force, and thereby the culprit of so many privacy threats. This suggestion completely misrepresents the issue. Technology has no autonomous will; it simply executes choices that are made by government, businesses, and individuals.

Because of technological advances (computers, increasing digital storage), we find ourselves in a world in which a tiny clerical error can propagate to devastating effects on a person's life (e.g. getting blacklisted, inability to apply for credit, etc.).

2. Central data storage *aids* rather than *threatens* privacy.

There is a widespread misconception that storing data in central "master databases" poses an additional threat to privacy. Actually the reverse is true. Data are stored for a purpose, and that "purpose" remains the same, the same need still exists, regardless of central or decentralized storage.

There are two major reasons why central storage is preferable. First, procedures to prevent unauthorized access to sensitive data are much more easily implemented and enforced on one location rather than ten. Secondly, privacy is only protected as well as the data quality permits. The most egregious examples of privacy violations in the past have invariably been accompanied by inaccuracies in data processing. These are much more readily restored in one location, rather than attempting to update ten different databases (with additional potential for yet more errors).

3. Privacy Protection Builds Trust

Protecting customer privacy is clearly in a company's best long term interest. It forms an integral part of a company's image and trust, its "brand value".

Unfortunately, because so many business leaders are focused on their next quarter's earnings, this simple truth doesn't always get the attention it deserves.

4. Identity Theft And Spam Are The Most Visible Violations Of Privacy Today

Although identity theft or takeover is (still?) largely an American problem, it is certainly on the rise in Europe as well (see tip #9, too). At an individual level this can have devastating consequences.

Spam is by far the most costly privacy violation of our time. It is estimated that 75-80% of all email messages are spam. That is over 45 Billion out of 60 Billion emails sent per day. Of these 45 Billion spam messages, 50-80% are sent by PC owners who have no awareness of taking part in these schemes, whose computers have been hijacked (so-called "zombie-PC's"). Interestingly, 200 known spammers are responsible for 80% of all spam (36 Billion spam messages per day). Talking about skewed distributions... J

5. Some Spam Trends

The use of zombie PC's and scanning of their email contacts has allowed spammers to circumvent spam-filters more often: the recipient's spam filter 'thinks' the message is from someone familiar and lets it through.

Another innovation is the use of graphic spam where the email message is technically in the form of a graphic image containing text (it *looks* just like a normal message). Few spam filters can detect this form of spam. In 2005 this was 1% of all spam, in 2006 so far 12% of spam is based on this method.

6. A Powerful Weapon Against Spam Is Ruled Out, And Knock-out

An effective method against spam was devised by Blue Security Inc. from California. Users connected to their system are added to a "do not spam" list. When they receive spam, Blue Security sends a warning. If that isn't effective, an electronic complaint is filed. The overload of complaints arriving at spammers' servers will effectively bring down their server.

However, the Coalition Against Unsolicited E-mail has rejected Blue Security's "denial-of-service" tactics. Deliberate attempts to bring down other people's websites are deemed illegal. In May of this year a Russian based spammer counter attacked, through the use of tens of thousands of zombie-PC's. They also threatened to flood Blue Security's customers with viruses, and that's how they were forced to close their operation.

7. Keep Your Email Lists 'Clean'

Real-time Blackhole Lists (RBL's) are a quasi-legal entity that gather IP addresses of known spammers. It's easier to get *on* than *off* those lists. One of the things they monitor is the number and kind of bounces that come off mass distribution of emails. ISP's do business with RBL's to filter spam at the network level.

This is one more reason why you want to keep your email list as clean as possible. Once your IP address is associated with too many bounces, you may be blacklisted or blocked. You keep your email list clean by setting (conservative) bounce rules, and then removing suspect email addresses. Of course you wanted to track your campaign results in as true a fashion as possible anyway, so there always was a good reason for maintaining your lists well.

8. How (un-)Desirable Are Unique Person Identifiers like Social Security Numbers?

There has been enormous debate about acceptance of national personal identification numbers. Until now, privacy activists have managed to prevent any numeric key from gaining wide acceptance. However, the question whether this serves our data protection purposes best is not quite so simple.

The reason for concern is typically that people don't like the 'anonymous' nature of a number. But databases never identify people, they identify keys, be they names or numbers. The technical advantage of a numeric key is that internal coding (check digits) can enhance reliability. And matches are always unambiguous. A downside is deemed that numeric keys will facilitate data merging (by criminals) and therefore enable record consolidation.

The tradeoff is whether the potential for abuse poses a greater threat than accidental merging errors by legitimate parties when more error prone keys like name/birthdate are used. 'False positive' matches have been known to cause considerable harm to innocent victims. Clearly the potential for similar (sounding) names to be inadvertently consolidated poses yet another data quality problem, creating harmless "victims" of well intended privacy protection.

9. Phishing Is On The Rise In Europe, Too

For quite a while, phishing wasn't much of a problem in Europe. This difference was caused, fundamentally, by our stricter privacy regulation. Data sharing can be done more freely in the US, and therefore criminals in Europe have a much harder time piecing together a victim's identity. Credit Bureau and state records are much less easily obtained in Europe.

In 2006, the number of phishing email attacks and the number of phony sites that aim to entice the consumer to leave sensitive details has almost doubled in Europe, compared to 2005. With these sensitive details (e.g. credit card, or personal data) identity takeover attacks are planned whereby the criminal will transfer funds in the name of the victim.

10. Data Protection Laws Have a Huge Impact On Companies' Ability To Market Their Products And Services

Database marketing has been, and continues to be, one of the key areas of focus of data protection legislation. The 1998 Data Protection Act applies to both *new* and *existing* customers. All need to be informed to what end their personal data are being, or will be used, and by whom. Non-obvious uses or disclosures should be properly described. For instance cross-marketing or list rental require prominent notification.

Compliance pertains to fair and lawful use. In laymen's terms this means that you need to be able to demonstrate that you actually need the data to run your business properly, and/or it is in the customer's best interest to do so. In particular "sensitive" data (a new, 1998 Schedule 3 definition), require explicit consent. These data are frequently used for credit scoring and targeting for instance. It is likely that currently many businesses are at risk of non-compliance claims.