www.rsm.nl/discovery

RSM DISCOVERY

Management Knowledge

23

3rd Quarter 2015

ROTTERDAM SCHOOL OF MANAGEMENT, ERASMUS UNIVERSITY

The business school that thinks and lives in the future

RSM Erasmus University

Handling threats to the validity of online data

By ***Petra Saskia Bayerl and Babak Akhgar***

Long gone are the days of online naivety where web users openly disclosed details about themselves and paid scant attention to cookies, trackers and privacy policies. Increased internet security awareness has doubled the pressure on businesses and public bodies to re-interpret the personal data they collect.

Today's web users do not necessarily have the time or the desire to trawl through the plethora of sites offering similar services or products. That said, many appreciate the customisation and personalisation of offers. Technologies now exist that can push such targeted content by matching recent online activities using sources such as cookies, trackers or user profiles. This approach is also in the commercial interest of the sites and services whose content is being pushed, as well as creating a million dollar business for companies offering personal data of potential customers.

The possibility of recuperating data available in the public domain (using so-called "open source intelligence" or OSINT techniques) means that businesses have a powerful tool enabling them to assess consumer product perceptions and purchase behaviours, track public opinions and consumer trends or measure customer loyalty whilst, at the same time, delivering content that matches people's (presumed) browsing and buying preferences. Of course not only private companies use such techniques. Governments and law enforcement agencies, for instance, also employ them to offer better services to their citizens as well as ensure their safety.

In this way, consumers – and the internet-using public more generally – become part of a "data-veillance ecosystem" and acquire "online data doubles" of themselves with which companies and governments try to predict with increasing accuracy the intentions and behaviours of the "real self" in order to sell their products, facilitate their administration or ensure a society's security. Internet users thus enter into a trade-off between using (often free) services from search engines to mobile games in the understanding that they pay for them with their personal data.

However, since the Snowden revelations internet users have become much wiser to this approach and the potential security and privacy pitfalls these practices harbour, which raises the question whether this increased awareness also has consequences for the acceptance of users for these data collection practices and thus for the business models relying on them.

The cost of privacy

The privacy policies of platforms such as Facebook have come in for considerable criticism in recent times. This particular case, as well as the repercussions for some Twitter users of their online declarations, has confirmed that there is no such thing as total online privacy. It is also an open secret that some HR personnel perform online searches and plough through social media to see how candidates present themselves beyond their CV and cover letter.

There are various techniques internet users can adopt to protect their personal data. Not accepting cookies or changing to services that guarantee higher privacy (from search engines and email providers to proxies or Tor-like networks) is one, while behaving with more restraint when publishing content on social media by very consciously choosing what to present in the spirit of "impression management" is another. A further approach is supplying false data when setting up online accounts or interacting with companies.

Handling threats to the validity of online data *(continued)*

By ***Petra Saskia Bayerl and Babak Akhgar***

Such behaviours – including avoiding certain platforms altogether – are viable options, but they can still mar users' online experiences and even mean exclusion from an increasingly vital arena of today's life. More and more web users are faced with a difficult choice between convenience and privacy, as are the businesses and organisations relying on user data for their sites and services.

Surveillance awareness

Our research illustrates that not only are web users becoming more sophisticated in protecting their privacy, but that site and service providers must also become more skilled in interpreting the data that they collect. The study in question addressed a population of 300 experienced web users primarily from the US, India, Canada, Croatia and Romania, with a balanced gender split and, in most cases, aged 40 or less. The line of questioning sought to establish and correlate two main issues – the perception and attitude of web users to online surveillance of their activity by private businesses and state organisations, and their opinion and practice of supplying false personal information, a trend that makes the collection and accurate interpretation of online data a major headache.

Awareness impacting attitude

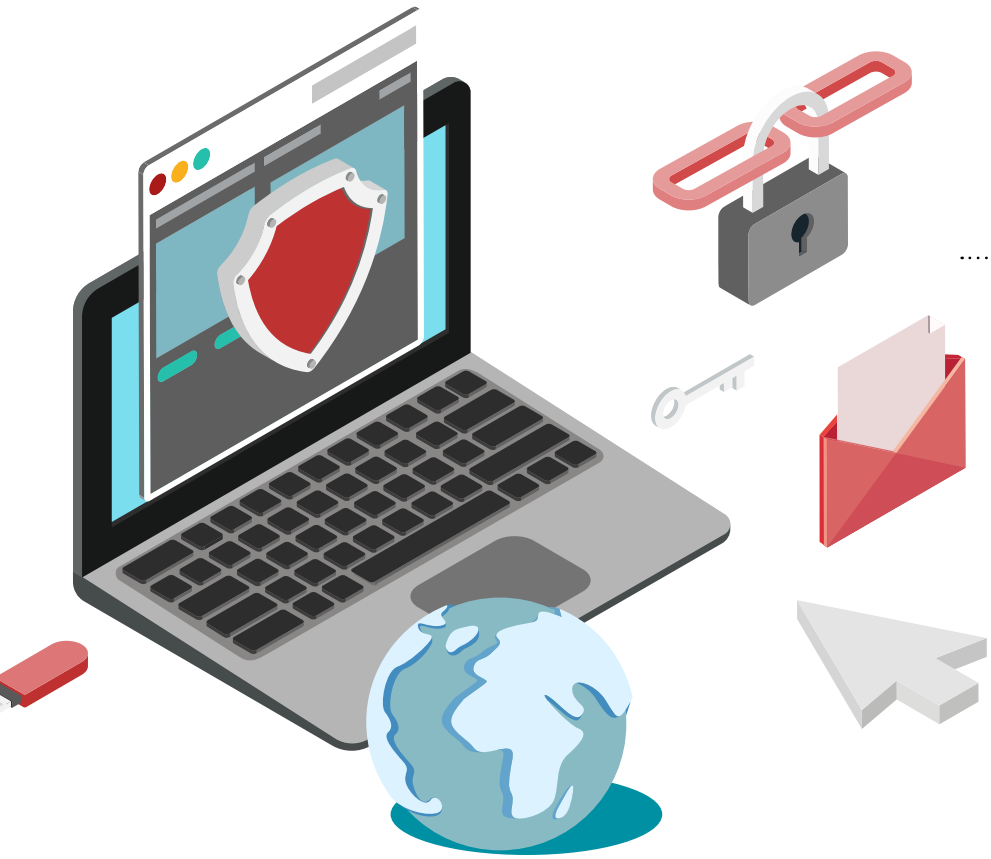
As one would expect, greater web experience generates greater sensitivity to the issue of online surveillance. The attitude to this practice varies according to who is doing the tracking and why. Stronger negative reactions emerge when the organisations in question are state agencies or governmental organisations, as opposed to private businesses. When carried out with legal objectives (ie, crime prevention) the activity is perceived more positively. In contrast, web users who feel their freedom of expression is threatened and/or the trustworthiness of their government is questionable regard the activity in a much dimmer light. At the same time, we found that for all users falsification of personal data is widespread – and widely accepted.

The key for businesses and state bodies is to understand how public perceptions translate into actual falsification of online personal data and the resultant obstacles this practice poses to accurate data interpretation. Interestingly, declaring the wrong gender and using an untruthful profile photo represent greater taboos than registering a pseudonym or bogus e-mail address, indicating that some information may pose greater challenges for verification than others.

Legitimising the act

Overall our study suggests that the key lies in whether data collection is seen as legitimised by its purpose – and thus in transparency regarding the act itself – how it is carried out, and why. No-one likes being caught out by fine print, so burying terms and conditions deep within an online policy statement will only irritate "internauts" and increases the chances of them providing false information or opting out altogether. Even worse are the instances in which problematic data collection practices are made public by third parties. Not only do they undermine the trust of the user community, but they also deter new users from joining.

Technical solutions exist for businesses that, in full awareness of the rising trend for data falsification, need to sort through the information collected with an even finer tooth comb. It has become quite common, for instance, to allow sign-up for a service only through a known account such as Facebook or to offer special benefits (discounts, gifts, etc) exclusively to users who sign in through such a known account. This exploits the possibility to map the social graph of web users across various online networks and perform classification and association mining. Further methods are trust score computational models and validity pattern mining. Yet, all these meth-



"Our research illustrates that not only are web users becoming more sophisticated in protecting their privacy, but that site and service providers must also become more skilled in interpreting the data that they collect."

ods are costly and do not address the growing pressures of online data collection on falsification tendencies and on user behaviours more generally. They operate on the symptom, instead of preventing the illness.

Onus on businesses

Awareness of data collection has never been higher, so the onus is well and truly upon businesses and organisations to be more candid and open in the use they make of any such information in order to legitimise this activity in the first place. Privacy has become a sales argument with which new services differentiate themselves from established competitors, offering for instance "no track" guarantees, fast decaying content or communication in exclusive communities. Users are thus increasingly provided with choices between more or less intrusive services, as well as ways to protect themselves from the prying eyes of companies and the government.

Whether and to what extent they use them depends on the perceived legitimacy of data collection and whether users can self-determine what they may or may not reveal. If there is no choice, falsification is an accepted option by many. Companies therefore need to develop a higher sensitivity of how their own privacy practices may push internet users into more or less open sharing of personal information and what type of changes they may have to expect. ■

This article draws its inspiration from the paper *Surveillance and Falsification Implications for Open Source intelligence Investigations*, written by Petra Saskia Bayerl and Babak Akhgar and published in *Communications of the ACM* Vol 58, No. 8, p62-69. DOI: http://dx.doi.org/10.1145/2699410

Petra Saskia Bayerl *is Assistant Professor for Technology and Organisational Behaviour and Programme Director Technology, Centre of Excellence in Public Safety Management, Rotterdam School of Management, Erasmus University Rotterdam.* EMAIL *pbayerl@rsm.nl*

Babak Akhgar *is Professor of Informatics and Director, Centre of Excellence in Terrorism, Resilience, Intelligence, and Organised Crime Research, Sheffield Hallam University, UK.* EMAIL *B.Akhgar@shu.ac.uk*