What if we couldn't be evil with personal data?

Angel investor John Taysom presents a radical new approach to protecting digital privacy while mining value from personal data

Have we paid too high a price for the wonders of the digital world? Like the fabled boiling frog, we've barely noticed how each new technological advance has worn away at our privacy. The trend has its roots in the origins of the Internet, which has sharing built in, whether we like it or not. The network's underlying protocols make messages easy to read and trace, and technologies like cookies, introduced to make the web experience more personalised and useful, mean our every online move can be tracked.

We contribute personally to the erosion of our privacy too, whenever we use an online service that doesn't cost money. It's no longer any secret that "if it's free, you're the product". Even a short burst of digital activity is enough to leave behind a detailed, personalised picture of yourself and your preferences. And at the same time as we are ever more pervasively tracked online, many of us, especially the young, willingly donate terabytes of personal data through social media.

Personal data: everlasting value

What's more, all of this data is now with us forever. Persistent storage is now the default. We like to talk of "digital footprints" but in reality our digital legacy will be much harder to erase. By 2011 it had become cheaper to store an item of data in perpetuity than to delete it. Our data has more in common with the pottery that fills museums of antiquity: almost indestructible, it will endure over centuries. Even if reduced to fragments it can be easily reassembled. Today's email thread is tomorrow's Ming fragment.

Organisations with a vested interest, which includes much of the tech industry, are fond of telling us that this is all good. Storing and sharing – of everything from status updates to biometrics – is the new norm. Today's digital commerce, especially advertising, depends heavily on personal data. 'Big Data' is the black gold of the tomorrow's digital economy: it's forecast to be a $16bn market by the end of 2014 and is also of colossal value for security, efficient government and the advancement of science. The "Internet of things" will soon add billions of connected devices to the Internet, all producing data that is likely to be valuable – and in many cases personal.

The vested interests have a point. If people were to stop feeding their data to the network, it risks damaging this value and choking off the digital economy now and in the future.

The road to *The Circle*?

Yet by continuing on our current course we may be heading for a surveillance dystopia of the sort imagined in Orwell's 1984 and more recently in Dave Eggers' novel The Circle. Privacy advocates are sounding ever-louder warnings. In recent comments, Edward Snowden characterised well-known Silicon Valley firms as "dangerous". From search habits to digitised genomic information, vast quantities of data are being placed in the hands of private corporations as well as governments. Consumers fear this too. The Ipsos 2014 Global Trends survey reveals strong concern worldwide about how companies use digital information. And according to TRUSTe, nearly three quarters of US Internet users were more worried about online privacy in early 2014 than in 2013.

For both business and the public, the urgent question is this. Can personal data be exploited for economic and social benefit without destroying privacy, killing e-commerce or even ushering in a totalitarian dystopia? Is Google's infamous "don't be evil" motto still achievable?

Privacy by default

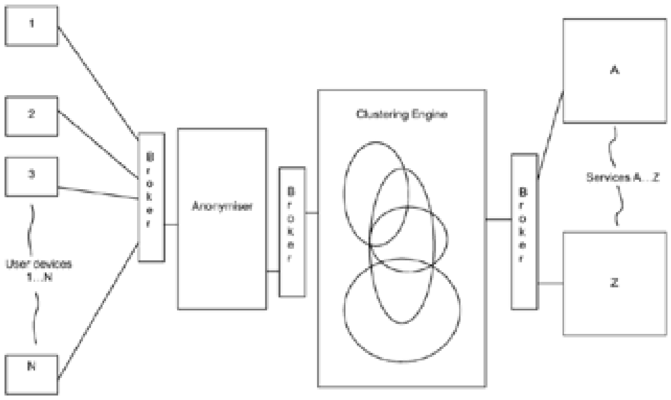
It is – but only if we can make personal data "private by default", while retaining the value of the data for those who need it. In effect, this would create a world where organisations "can't be evil".

The important insight is this: what's interesting to advertisers, genomic researchers and the security services is not our identity, but our similarities. With a way to 'cluster' data – like purchase behaviour, location, even the setting of your smart thermostat – together with that of others, information remains valuable but can be rendered untraceable to an individual. This way, it's possible to preserve privacy while at the same time maintaining data utility. It's not just about keeping advertisers happy: this could unleash the potential of data such as GIS and genomic information that is currently hard to exploit because of privacy concerns.

Virtual hoodie

Think of the model as a "virtual hoodie" that conceals your identity but not your characteristics. Privacy-protected tracking works by generating real-time clusters of individuals with similar characteristics, rather like virtual zip codes. But in this case, targeting can be more fine-grained yet still not invade privacy.

The approach depends on three innovations. The first is a protected online repository of personal information; the second, a not-for-profit, non-governmental institution, akin to ICANN, to administer it. Finally, a clustering algorithm ensures that useful personal information is shared, but only in a way that conceals individual identity.



The "private by default" model puts an information broker and clustering engine between the user and the service they are using, so users are only identified as part of a group of similar individuals.

Much to gain, nothing to lose

Could this happen? The technology solution has already been patented and discussions have already taken place with policymakers. There's no need for government oversight or new regulations. Most importantly, both sides of the transparency vs. privacy debate are winners. It's a prize worth achieving because it could also catalyse existing and nascent industries that depend on data: online advertising, genomics, security, telemetrics and the Internet of Things. And a federated approach to data sharing could also reconcile the privacy standoff between the US and EU.

Ultimately, privacy underpins autonomy, the basis for freedom. We can't carry on collected and exploiting personal data without damaging freedom itself. Business has much to gain and nothing to lose if it can't be evil.

john@taysom.com