

Closing remarks: Data-driven Innovation, Ethics and Technology

During this Forum, we have heard several examples of the use of big data for emergency response, such as earthquake recovery and the many applications of Ushahidi. No-one could sensibly argue against these as legitimate and valuable uses of big data.

But many other potential uses have also been described: autonomous vehicles or even weapon systems; machine learning systems, behavioural prediction, and so on. And we need to be clear: for all the *possible* beneficial uses of big data, the number one use to which it is put today is for targeted marketing.

Today's supposedly "free" Internet services are funded by the monetization of personal data, in ways that are outside the control and awareness of the data subject. So when it comes to potential future uses of big data, I want to make sure we are asking the right questions. After all, we have been here before: you will have heard the phrase "personal data is the new oil", and it is a phrase which appears in the Synthesis Report. But what has experience taught us about oil?

Its use brings economic growth, productivity benefits and the personal convenience of global mobility, yes.

But:

- Its extraction is neither ethically nor ecologically neutral.
- It is toxic, and can also be flammable and explosive.
- Its processing, storage and use generate risk, and harmful by-products which are changing our planet's climate, possibly irreversibly.

In other words, if someone first proposed oil, today, as the new answer to the World's economic problems, we would have some serious questions to ask them about the wisdom of their proposal. I'm sure you can see parallels between what I have said about oil and the drawbacks of collecting, processing and storing personal data.

Above all, I want to draw the distinction between what is technically possible and what is socially desirable. As someone put it clearly yesterday, "Trust Is Social Capital". Information technology now reaches into our lives to such an extent that the technical, societal and economic factors cannot be disentangled. To contemplate one is to contemplate the others.

Technical innovation often asks economic and social questions to which technology does not have the answer. Let me give an example, to make this topic less abstract.

The use of location data (for instance, collected from a mobile handset or intelligent domestic objects) may begin to give early clues about the onset of Alzheimer's. At first sight, this looks like one of those beneficial uses of big data which we should adopt without hesitation.

Before big data, one can imagine that concerns about early signs of dementia might surface through a conversation between the individual and his or her doctor, and would be confidential between the two of them.

But in the big data context, the overwhelming probability is that incipient Alzheimer's would first become known to a third party; quite possibly a third party with no professional or medical relationship to the individual. After all, your device manufacturer is not your doctor. The question then is: what should be done with that information?

Should it be communicated to the individual's primary healthcare provider? And if so, is that a state healthcare professional or a commercial one? Or perhaps it is an insurance company? That may depend on which country we're talking about, and in each case, the third party's societal and economic incentives may be different.

Should it be communicated to the individual's family or partner? Bearing in mind that Alzheimer's can lead to a loss of mental capability, there may come a point at which the individual is no longer legally competent. At what point does the data mean we should *not* tell the individual, but should deal only with someone to whom authority has been delegated on their behalf?

I describe this example because the problem of delegated authorisation is already known to be a difficult one for technology to solve. It is difficult to solve, even if what you want is a digital authorisation for your neighbour to collect a parcel from the post office on your behalf. And yet here, we are talking about what to do about someone with the initial signs of dementia. Not collecting a parcel.

Let me give one more brief example, this time concerning autonomous vehicles. At some point, a driverless car will be faced with a choice between hitting a pedestrian, or taking avoiding action which will result in injury to the vehicle's occupant. Who will take responsibility for the outcome of that choice? In terms of agency, it would seem almost as unfair to blame the occupant of the vehicle as it would to blame the pedestrian. But I would be willing to bet that, on the purchase agreement signed by the owner, the small print will say that no liability remains with the developers of the vehicle's hardware or software.

As I say: technical solutions can raise societal and economic issues to which technology does not have the answer. We need to be aware of that, as we rush enthusiastically towards the next exciting technical innovation.

With all this in mind, I have three closing recommendations, in which I include views expressed by my colleagues in the Internet Technical Advisory Committee:

1 – The current GDP growth assumptions for the contribution of the Internet to Data Driven Innovation are based on the fundamental premise of a single Internet. So, for example, research clearly indicates that a fragmented DNS infrastructure would significantly impact future growth. We need to construct a governance regime which discourages governments from the idea that cybersecurity or other policy issues are best addressed by "closing off parts of the Internet" – because that approach threatens the fundamental premise and therefore also the forecasts of future growth.

2 - There is a general lack of consensus on what constitutes personally-identifiable information. This has a knock-on effect on practical measures to combat cyber crime, such as cross-border data sharing for law enforcement, and the definition of "Internet-appropriate" cross-jurisdictional warrants. For instance, as previous speakers have noted, metadata and inference data can be just as "personally identifiable" as a name and address, but are treated very differently in regulatory terms.

3 - We need to ask, much more often, the question "*should* we do this", not just "*can* we do this". Too often, the present approach of governance, risk and compliance (GRC) becomes an exercise in ticking the boxes on a compliance checklist, and the result is to minimise the liability of the organisation, not to protect the interests of the data subject.

The Internet Society would like to see a focus on ethical data-handling [1]. Organisations should ask themselves:

- Does this use of data genuinely reflect the interests of the data subject as well as the interests of the organisation?
- Is there transparency and accountability in its collection, sharing and use?
- Would this use of data come as a surprise or a shock to the individual concerned?
- When the organisation faces a choice about what to do with data, which option represents the greatest fairness, transparency and accountability?

In this way, perhaps we can move towards a more ethical balance between the interests of the data controller and the individual.

Robin Wilton
Technical Outreach Director - Identity and Privacy
Internet Society

Tokyo - 3 Oct 2014

[1] Ethical Data-Handling (white paper): <https://www.internetsociety.org/sites/default/files/Ethical%20Data-handling%20-%20v2.0.pdf>

Internet Society
Chêne-Boulogne, 15

Switzerland
Tel: +41 22 807 1444
Fax: +41 22 807 1445
www.internetsociety.org

1775 Wiehle Ave.
Suite 201
Reston, VA 20190
USA
Tel: +1 703 439 2120
Fax: +1 703 326 9881
Email: info@isoc.org