Privacy and Autonomy in Ambient Intelligence

Caspar Bowden

(privacy advocate and consultant)

2002-2011 Chief Privacy Adviser, Microsoft
1998-2002 Foundation for Information Policy Research UK

LAAS ADREAM inauguration

5th July 2012

summary

- EU DP regime has serious conceptual flaws
  - no “quick fix” from new EU Regulation
- “Privacy engineering”: new interdisciplinary field
  - PETs can do “magic”: but will users understand?
  - research “gap” in Human Computer Interaction
- ambient systems: unique challenges ahead
  - needs basic research AND engineering
- future competitive opportunity for Europe
  - largest market with strong uniform privacy laws

Non-technical problems of privacy

- DPAs don't understand computer science
  - ~2500 DPA officials EU: <1% with comp.sci
- “consent” not a technical issue (or “formality”)
  - not lawful unless freely given => reasonable choices
- business lobbying crippled 1995 EU Directive
  - allowed US Internet companies to flout EU law
- perverse effects from behavioural economics
  - “frog-boiling” intuition is wrong!
  - users “punish” good privacy design

What is “identifiable”? EU DP Directive 1995

- Article 2
  - *(a) 'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an **identifiable** person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity;*

The central question: Identifiable by whom?

- Recital 26
  - *Whereas the principles of protection must apply to any information concerning an identified or identifiable person; whereas, to determine whether a person is identifiable, account should be taken of all the **means likely reasonably to be used** either by the controller **or by any other person** to identify the said person; whereas the' principles of protection shall **not apply to data rendered anonymous in such a way that the data subject is no longer identifiable**.*

- Q. “means” – must refer to legal methods as well as statistical methods?
  - Otherwise passport/health-insurance/credit-card numbers would not be personal data (in themselves)
- Q. “means likely reasonably to be used” – must refer to the *method* not the *individual person*?
  - Because otherwise implies a person can be deprived of ECHR Art.8 privacy rights because in a minority!
- Q. “or by any other person” – seems to imply that the following rider “whereas…rendered anonymous” has to be understood as anonymous with respect to “any other person”, not merely with respect to the putative data controller?
- Conclusion: if identification might likely reasonably occur by lawful process or statistical means (but even if only happens to a minority of individuals) it **is personal data**.

2012: proposed new EU DP **Regulation** (uniform over EU)

- (some) govts. already trying to limit to “easily” identifiable data

RFID policy 2005:
EU Commission ignored DPAs

Working document on data protection issues related to RFID technology (WP 105) Jan 2005

3.3. Use of RFID to track without "traditional" identifiers being available

- For example, there is the possibility for a chain grocery store to give out tagged devices to customers (e.g., like tokens) enabling the operation of shopping carts, which customers re-use each time they visit the store. Such a mechanism would permit the store to set up a file using the identification number stored in the tagged device enabling it to monitor which products an individual (identified by the token) purchases, how often such products are used and in which of the chain grocery stores the consumer buys them. The store could make inferred assumptions about an individual's income, health, lifestyle, buying habits etc. This information could be used for various decision making, such as marketing, purposes or even for dynamic pricing. Since the device would identify the individual each time he/she entered the store, the consumer could be marketed to in the light of the recorded consumer habits. In addition to the store being able to collect the above information, a third party could potentially also obtain such information**. In this way, various decisions could be made about that identified individual without his or her informed consent**. As it happens, with the **use of cookies in the on-line environment, even if the individual is not immediately and directly identified at the item information level, he can be identified at an associative level because of the possibility of identifying him without difficulty via the large mass of information surrounding him or stored about him**. Furthermore, the data collected from him can influence the way in which that person is treated or evaluated. This RFID use also carries serious data protection implications.
- A further example could be where the use of RFID tags can lead to the processing of personal data, even when RFID technology does not involve the use of other explicit identifiers. Take the hypothesis where person Z walks into Shop C with a bag of RFID tagged products from Shops A & B. Shop C scans his bag and the products in it (more likely a jumble of numbers) are revealed. **Shop C keeps a record of the numbers. When person Z returns to the shop the next day, he is rescanned. Product Y, that was scanned yesterday, is revealed today** – the number is for the watch he always wears. Shop C sets up a file using the number of product Y as a 'key'. This allows them to track when Person Z enters their shop, using the RFID number of his watch as a reference number for him. **This allows shop C to set up a profile of Person Z (whose name they don't know)** and to track what he has in his shopping bag on subsequent visits to Shop C. By doing this, **Store C is processing personal data and data protection law will apply**.

De-anonymization/Re-identification

- “Anonymous” data isn't
  - personal/non-personal distinction is (mostly) untenable
- Shmatikov/Narayanan “Netflix” paper 2007
  - other results on social networks (also Dwork)
  - cheap, easy, general: no supercomputer needed
- rise of social networks: Facebook/Linked-In/Twitter
  - world is awash with social graph auxiliary data
- Intersection attacks very powerful (especially location)
  - “granularity” doesn't work
  - daily commuting: endpoints identify 95% people
- Smart grid/meters: can infer intimate details of home life
  - possible to use advanced crypto so no personal data leaves meter!

new computer science of privacy

- PET Symposium (2012 in Vigo: registration still open!)
  - PET Award 2003-
  - “Digital Privacy: Theory, Technologies, and Practices” 2007 Acquisti et al. (best survey book for introduction)
  - Tor (onion-routing), robust theory of communications anonymity
- EU Framework research: PRIME, PrimeLife
  - “private credentials”: ABC4Trust, U-Prove
- Privacy languages:
  - EPAL, XACML, APPEL, P3P (no semantic web)
    - but no European privacy rights intrinsic in semantics :-(
- From “data minimization” to risk minimization
  - Differential Privacy (Dwork), homomorphic encryption (Gentry)

Ambient systems: problem statement

- in ubiquitous computing, many nodes communicate autonomously – dependability, resilience, “safety”
- the (benevolent) goal is to serve human needs and use resources efficiently - but risks of alienating people
- maybe the system can represent user's intentions, and anticipate wishes without explicit control ?
- ...then system must learn and model user's preference
  - => more the system knows, more privacy risks !
  - state of AI/machine-learning – unlike 1980s expert systems, hard to predict or explain ethical reasoning or consequences

The Private Sphere

- intimate, reflective states of consciousness
  - Josh Harris: “We Live in Public”: no privacy is pathological
    - “First Person” interview Errol Morris
- ambient “private sphere” processed by many CPUs
  - little data can cause a lot of damage
- augmented reality and virtual worlds as UI ?
  - ethics of “do you see what I see?”
- user-centric vs. system-centric design
  - how many persona/pseudonyms are “normal” ?
  - story of Microsoft U-Prove and corporate pathologies
- how hard should designers try for useability ?
  - tragedy of privacy economics

limits of “control” in Ambient

- perceiving and representing intention is hard
  - machines can learn, but can they explain?
- designing for vulnerable users
  - delegation is essential, but still ethical problems
- who writes the rules?
  - (most) users won't write explicit rules
  - sysadmin ? “trust” is not the point. Autonomy
- where are privacy computations done ?
  - Cloud = outside-security-perimeter?

is there a button to stop the system “pushing-my-buttons” ?

- Ambient Privacy Paradox:
  - system can only adjust automatically if knows user preferences
  - can only control if insight into what system believes
- How much do you want system to know ?
  - bio-sensors to adjust your mood?
    - “Don't want to relax – have a deadline – make me stressed!”
  - The system's representation of “private sphere” is (maybe) even more private than private sphere
    - What if hacked ?
    - What if system reset erases years of “training”?

Need basic research in privacy science/engineering

- Privacy-by-design in new EU Regulation
  - fines can be up to 1% or 2% **<u>global turnover</u>**
- social networks erode privacy, but young people still want/need control (danah boyd)
  - privacy will become *more* valuable as becomes more scarce
- privacy and security interdependent
  - technical InfoSec must be intelligible, socially explicable

Conclusions

- designers should be aware of regulation “real-politik”
  - just because “lawful” doesn't mean ethical !
- privacy engineering needs holistic analysis
  - properties do not “compose” (unlike security)
- A new legal principle: privacy-about-privacy-rights?
  - no reasons to intrude on use of privacy rights?
  - e.g. police monitor frequency of “subject access” ?
    - cannot make exercise of a human right “ suspicious”
      - else undermines the right!
- Privacy is interdisciplinary: but where can one learn relevant law, comp.sci, economics, psychology, pol.sci ?