Expert Policy Panel: Competitiveness, privacy and marginalization challenges of big data research

Indicative questions

1. Increased attention being paid by companies to insights drawn from big data, especially about customer behavior, suggests that it is a critical competitive resource. Do you agree?
2. If yes, will access to big data, especially to data sets with comprehensive coverage, become a source of competitive advantage or disadvantage?
3. “Inform and consent” is the conventional paradigm for safeguarding privacy. It emerged in the context of regulating credit information.
	1. Is it possible to apply this paradigm to big data, especially transaction-generated data such as call-detail records (CDRs) or visitor-location registry (VLR) data (for definitions, see Annex 1)?
	2. Specifying the use is an integral element of “Inform & consent.” Is it possible to specify the uses of big data?
	3. Will the application of I&C to big data result in this competitive resource being available only within the company and third-party, upstream firms being shut out of access? Will this tip the scales toward large conglomerates and discriminate against start-ups?
4. Privacy is a concept that is difficult to define (see Annex 2). In actual practice it is embedded in a rights framework. Rights are, for the most part, individual. But many insights derived from behavioral big data or transaction-generated data are about collectivities.
	1. Can an inherently individual concept such as the right to privacy be extended to collectivities, in the form of a “right to collective privacy”? An example of “collective privacy” is given as Annex 3.
		1. If the “right to collective privacy” is recognized as worthy of protection, will it be the end of social science as we know it?
		2. If the “right to collective privacy” is recognized as worthy of protection, will it be the end of targeted policy interventions?
5. Those who advocate strong privacy protection may imply that being insulated from data collection is an ideal condition. But what are the consequences of being excluded from data streams that influence resource allocation? See Annex 4 for two illustrations.
	1. As the world moves toward greater reliance on data analytics, will we see a new divide emerging between those who are visible to the data analysts and those who are not?

# Annex 1: CDR and VLR data

Mobile Network Big Data include call-detail records (CDRs) generated when calls and texts are sent/received, Internet is used and prepaid value is loaded, and visitor-location registry (VLR) data that are generated when handsets “tell” base transceiver stations (BTS) that they are in the coverage areas. CDRs require some action by the user. VLRs do not. VLR data are generated independently of the user as long as the phone has power. CDRs, which include data elements such as calling-party number, called-party number, the BTS where the call originated, time of call, duration and information about the device, are used for billing purposes. Therefore, the data is stored for some time. VLR data are larger in volume and tend to be written-over. Other than some data necessary for network management such as those on load factors, VLR data are not routinely retained.

# Annex 2: On Privacy

Many aspects of identity and self are covered by the common-sense understanding of privacy. Privacy, as commonly understood, “is a sweeping concept, encompassing (among other things) freedom of thought, control over one’s body, solitude in one’s home, control over personal information, freedom from surveillance, protection of one’s reputation, and protection from searches and interrogations.”[[1]](#footnote-1) Attempts to define it in terms of boundary control by individuals[[2]](#footnote-2) are difficult to translate into practical policy. For example, where are the bright-line boundaries of what an individual has authority over in the case of data generated as a by-product of a transaction?

Solove argues that privacy as an abstract concept is difficult to pin down, since it “involves a cluster of protections against a group of different but related problems.”[[3]](#footnote-3) He concludes correctly, that the focus should be shifted away from defining privacy, to addressing privacy problems (or harms). He proposes 16 privacy problems, grouped into four general types: Information collection (comprising surveillance and interrogation); information processing (comprising aggregation, identification, insecurity, secondary use and exclusion); information dissemination (comprising breach of confidentiality, disclosure, exposure, increased accessibility, blackmail, appropriation and distortion); and invasion (intrusion and decisional interference).

# Annex 3: Collective privacy

It is widely believed that there is greater consumption of adult or pornographic entertainment when conventions attended large numbers of Christian Evangelicals are held at US hotels.[[4]](#footnote-4) Whether true or false, this perception is harmful to the image of Christian Evangelicals.

Let us assume that there is hard evidence to substantiate the above claim. For that, it would not be necessary for the hotels to release the video viewing records of individuals, violating the provisions of the US Video Privacy Protection Act of 1988. Instead, they could simply provide the aggregate use records by title or category of videos. With this information, one could observe the peaks and valleys of adult entertainment use by date in a specific hotel or hotels. It may be mildly interesting, but not newsworthy.

However, if the information on daily consumption of adult entertainment is correlated with the numbers of hotel guests attending specific conventions, be they atheist or evangelical, the story begins to become interesting. If one can establish a consistent correlation between increased consumption of adult entertainment over a period of time and with particular kinds of conventions, especially if the viewing of such movies is contrary to the public positions taken by the convention organizers, the story becomes truly interesting. It could, ostensibly, be damaging to a group that is hostile to depictions of sexual behavior in entertainment.

This is an example of a breach of collective or group privacy, as commonly understood. The simple aggregation of individual video rental records does not constitute the breach; it is the combination of that data with data identifying the group. Here too, the conceived harm is connected to identification of the group.

# Annex 4: Examples of marginalization

**Boston’s Street Bump app**

The City of Boston makes available an app called Street Bump that can be downloaded to smartphones. Any citizen can place the smartphone in a holder in a car and press one button to start the app at the beginning of a journey. No calls would be taken during the journey. The accelerometer of the smartphone collects data that has been proven to be effective in identifying pot holes and speed bumps. At the end, another button is pressed and the collected data including the GPS coordinates of the starting and ending points are sent to City Hall. Using algorithms the bumps that should be there and those that should not be there are identified and the latter get routed into the work order system for repairs.

The assumption is that smartphones are ubiquitous in Boston. What if a similar crowdsourced big-data application is deployed in a city which has less than 10 percent smartphone users?

**Prohibitions against transborder flows of credit information**

Anyone who has moved to a new country and tried to obtain services that require a credit rating will understand the consequences of exclusion. The earliest data protection laws were enacted in response to the rise of credit cards and credit reporting.[[5]](#footnote-5) They were intended to safeguard against problems of data aggregation, identification, insecurity, secondary use, etc. But an unintended consequence of rules against credit reports being sent across national borders was that individuals who crossed those borders were denied services or had to make do with more expensive options until they built up credit histories in the new country.

1. Solove, D.J. (2008). *Understanding privacy*. Cambridge MA: Harvard University Press, p. 1. [↑](#footnote-ref-1)
2. E.g., Samarajiva, R. (1994). Privacy in electronic public space, *Canadian Journal of Communication*, 19(1): 90. [↑](#footnote-ref-2)
3. Solove, D.J. (2008). *Understanding privacy*. Cambridge MA: Harvard University Press, p. 174. [↑](#footnote-ref-3)
4. <http://gospeldrivenchurch.blogspot.com/2011/03/what-you-do-in-your-hotel-room-gives.html>. This site is sympathetic to Christians and hostile to adult entertainment. [↑](#footnote-ref-4)
5. Rule, J.; McAdam, D.; Stearns, L.; Uglow, D. (1980). *The politics of privacy: Planning for personal data systems as powerful technologies*. New York: Elsevier, ch. 6. [↑](#footnote-ref-5)