Data retention in the European Union

Leistert, Oliver

Published in:
International Journal of Communication

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

International Journal of Communication 2 (2008), 925-935 1932-8036/20080925

Data Retention in the European Union: When a Call Returns

OLIVER LEISTERT

University of Paderborn

Retained (via digital storage) metadata of telecommunications acts change and transform into the content of something else: a surveillance program. Originating from telecommunications as a protocol necessity, the metadata is fed into a data space that freezes and manipulates the time axis. This measurement of post-9/11 governing is only one item in an assemblage of surveillance technologies that are not watching, in the manner of traditional CCTV, but processing the population under observation. Since data processing is a more recent, counterintuitive, and still relatively opaque principle, its capacities are not adequately understood, nor are they established in the general understanding. The concept of a data space that provides movement within and between data described here illustrates the powers of data retention in an imaginable way.

Introduction

This article's subject is a proposal for how to attain a more intuitive point of view of the seldom recognised, immense power of data retention in a supranational information space like the EU. A problem of massive data collection in huge databases with highly sophisticated information technologies, such as data mining, lies in its imperceptibility. While an analog archive is impressive simply by its sensual perceptibility, digital data collections do not matter at all on the phenological side.

By providing more intuitive concepts around the metaphor of *data space,* this article's main intention is to demonstrate the powers of data retention from *within* the technologies' logic as an immanent approach. Further, it suggests that the so-called metadata of communication acts cannot be regarded as such until they are being treated as content data on a different level — the level of data retention. The article's cause is the Data Retention Directive of the European Union, which forces the production of an unequalled decentralized accumulation of all metadata of communication acts (e.g., call detail records of telephony and Internet traffic) within the EU.

Oliver Leistert: **oleist@zeromail.org**
Date submitted: 2008-02-28

The Directive in a Nutshell

The European Union Directive on data retention,[1] though less than 10 pages long, is invested with considerable authority. It directs the member states to pass a law compelling each provider of telecommunications services to retain *traffic* and *location* data for at least the past six, and at most, the last 24 months. As stated in the first sentence of Article 1:

> This Directive aims to harmonise Member States' provisions concerning the obligations of the providers of publicly available electronic communications services or of public communications networks with respect to the retention of certain data which are generated or processed by them, in order to ensure that the data are available for the purpose of the investigation, detection, and prosecution of serious crime, as defined by each Member State in its national law. (2006/24/EC: 56)

Certain data here means *traffic* data and *location* data, and as defined here, is data generated by or during an act of telecommunications with a mobile phone, a landline, or via the Internet, minus the "content." These inquiries around data ask who, when, where, with whom, how long, and so forth — but do not ask about the nature of the communication. The data generated during *unsuccessful* acts of telecommunication is also similarly analysed.

To "harmonise" means to implement technical standards of retention, and to do so for data access from anywhere in the EU.[2] The data that one profiler gets from a member country shall technically be compatible with the data s/he obtains from any other member state.

Unquestionably, the Directive rests on a differentiation between traffic/location data and content data. The retention of *all* data generated during an act of telecommunication might not fit with Directive 2002/58/EC on data processing and privacy of July 12, 2002, and other fundamental human rights.[3]

---

1 European Union directive 2006/24/EC of March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communication networks and amending directive 2002/58/EC, hereafter: 2006/24/EC.

2 It is worth mentioning that this directive has passed as a so-called First Pillar directive using the single market power of the European Union, and not as a Third Pillar directive pursuant to the Union's power fighting crimes. "Once the choice was made to go ahead with the Directive as a First Pillar initiative, the Commission and the Council took the position that, legally speaking, the Directive could not regulate police access to communications data. Anything having to do with the police was strictly Third Pillar" (Bignami, 2007, p. 12).

3 If the actual implementation does so, and if it complies with national laws on privacy, is not of concern here, and neither is the discourse on data retention and citizen rights. For a historical discussion of data retention in Europe and Canada, see Warner, 2005.

Traffic Data and Content Data

This differentiation between data that contains the structural components of communication and data that relates to content is, first of all, technically inspired, and thus indicates the possibility that the Directive is produced with this technical differentiation in mind. Traffic data consists only of the information needed to technically initiate, sustain, and terminate an act of communication. However, as the Directive aims at "the investigation, detection, and prosecution of serious crime," the relevant data is divided into the following symbolic subcategories:

- data necessary to trace and identify the source of a communication;
- data necessary to identify the destination of a communication;
- data necessary to identify the date, time, and duration of a communication;
- data necessary to identify the type of communication;
- data necessary to identify users' communication equipment or what purports to be their equipment; and
- data necessary to identify the location of mobile communication equipment.

Obviously, this subcategorisation of traffic data is not *inherent* to the data (as it is produced through the cooperation of user, device, and infrastructure), but this subcategorisation represents a scheme of the profiler's questions. In addition, the term *communication* itself is problematic here. Article 6, Sentence 2, of the Directive states: "No data revealing the content of the communication may be retained pursuant to this Directive." But the person calling the crisis line is not ordering pizza. There are undeniably *semantic elements* in the mass of so-called traffic data. Data identifying the person calling the crisis line unavoidably reveals content (by virtue of the *nature* of the call.) This is one of the reasons why such data is protected under privacy legislation in the European Union.

Two Symbolic Worlds

The symbol-processing machines called computers that enable telecommunications services, and the symbol-"processing" (or more accurately, "recognising") humans trying to communicate via these media are not equivalent to each other in terms of hermeneutic and cognitive function. The realm of the symbolic mediates the trajectories of the real and the imaginary *for* and *between* humans. The symbolic — constituted by signs and by modes of signification — is the basis of any human communication. Machines also process symbols, but neutrally (i.e., neither in reference to the outside nor to the inside).[4] They don't rely on a world representation in their minds.[5] Computers are radically autistic (Krämer, 1992, p. 339) in

---

4 For a precise discussion of the common misunderstanding of computers as enhancements of humans and the anthropocentric (mental) case, see Tholen, 1994.

5 A lesson any "artificial intelligence" research had to learn.

terms of human characteristics. Their only concern is whether "it" (the encoded datum) is computable or not.[6]

Computer-processed data does not relate to anything outside the machine. The caller's ID is a device ID. The profiler's assumption that the device is equal to, or identical with, or identifiable with its user is an obvious *pragmatic* reduction. It serves here as a metaphor for the appropriation of technical data for profiling issues, because it may be argued that this is not the same *kind* of data anymore: Traffic data is *transformed* into something else in order to lend itself to specific uses, including those of surveillance, by the profiler.

Interfaces as Gates between Symbolic Worlds

Interfaces connect computers with the outside world, and vice versa. It is via interfaces that input and output can be processed. A computer without any interface is both a paradox and an impossibility.

Data entered into computers for a telecommunication act are to be divided into "outer machinic" and "inner machinic." The ubiquitous Internet Protocol (IP) is an apt example: Domain names are outer machinic data as they relate to the outer world, while IPs are inner-machinic addresses of the Internet, without reference to its outside. A domain name server does the necessary translation between the two modes.

In this analogy, IPs are traffic data, while domain names are content data. This neatly describes the conflict in a nutshell. IPs do not represent anything but a numerical address of a specific machine. Domain names do not represent a specific machine, but signify codes created by and comprehensible to humans, such as "fbi.gov".

Retaining traffic data *shifts* the *address space of meanings* from machinic to human. This is precisely why ethical problems may occur. Computers process symbols regardless of their meaning, as long as they are operational. Human beings process symbols regardless of their technical viability, as long as they are meaningful. Two ontologically distinct worlds collide here[7] as data from one world is fed into the circuits of the other.

---

6 The radical difference between computers and human beings is constantly being blurred by anthropocentric thinking of technology. Viability, as the leading paradigm in science and technology, also pretends connections between machines and the outside world that do not exist (Winkler, 2004, pp. 226-230).

7 It is the blindness of the machines that in the first place renders possible the signifying human work. Data collection, processing, and mining can only be done if the machines do not interfere with opinions about the meaning of the data. Generally, there are two ways out of this dilemma: enter ethics into machines or disallow certain machinic operations.

While the traffic data is generated during a telecommunication act, its signification for humans and outside of the processing computers belonging to the telecommunications infrastructure itself is produced *by its retention*, via standardised access and, later, its reference to people's names.

Taking this shift seriously, one may speak here of "new" data and reject the notion of the common identity of traffic data and retained data. The retained data has no operational meaning anymore, but it is transformed into the symbolic: It now *represents* the movements and telecommunication acts of people.

As the traffic data is produced "automatically" by the communications technology itself as a working necessity, the production of the profiler's data is very economic. It is not even necessary to introduce new hardware or sensors into the existent infrastructure.

By retaining the data for the purpose of crime investigation, a significant shift transforms[8] the prior technically necessary data into data that now has a meaning for humans, and so is strictly speaking some sort of *content* data. The process of retention itself inherently supports the transformation of data: Now, it is archived on some dedicated storage media in a dedicated storage form, easily accessible and searchable.

This transformation of traffic and location data into content data cannot be described within the logic of the act of communication itself, as this act is outperformed by machinic operations. The database *containing metadata* is a newly generated object, produced by specific algorithmic operations and strategic settings.

A New World

A database is a "collection of data or information organised for rapid search and retrieval."[9] In this case, the database is needed to ensure the persistence of the ephemeral by generating durable data sets. These data sets consist of successful or unsuccessful acts of telecommunication registered and executed in the telecommunication infrastructure within the geographical space of the European Union. Data thus transformed does not constitute an object that belongs to the telecommunication infrastructure itself; rather, it is a materialisation that rests on a massive time-axis manipulation. Essentially parasitic, it negates existent chronology and transforms expired data logs into valid current information.[10] It generates an ahistorical time window of 6 to 24 months and consists of symbolic representations of a compressed

8 As Bruno Latour puts it, "There is only transformation. Information as something that will be carried through space and time, without deformation, is a complete myth [. . .]. From the same bytes, in terms of "abstract encoding," the output you get is entirely different, depending on the medium you use." (Latour, 2004, p. 154).

9 http://www.britannica.com/ebc/article-9362288

10 Winkler calls such as state a co-presence of past and present (Winkler, 1997, p. 175).

space-time manifold through data doubles — or "dividuals," as Deleuze (1992) puts it — preserved for profiling and crime prosecution. As a form of decentralised surveillance architecture, the data retention database infrastructure is very robust. This is because metadata is always processed from two sides, that of the sender as well as the receiver.

Since the time axis is thus deconstructed and then reorganised by the storage operation, profilers enter a hyperreal/surreal world of replicated identity. This strategic copy is navigable through space and time. User locations and communications now can be reconfigured into feedback loops and transmitted back and forth arbitrarily.

Twofold is a Ticket

The convention to retain data on both sides of the communication renders possible data-travel along these electronic meridians by following communication patterns of constructed groups. Starting with database d1 and data double a1' of communication act a1 that connects device A and B, the travel continues to database d2,[11] containing the data double a1'', which represents the connection of A and B once again, but from the reversed perspective. Any connection of B with C is now within reach, followed by C with D, and so forth.

This doubling of representation enables the easy construction of groups whose members have nothing in common in real life. Where communication acts of one single person were analysed, the topology was simple and compact: a star with the person's device in the middle as the connecting point with the rest. What can be called a "crawling" topology enabled via the storage of the two perspectives of communication act is potentially endless and self-generating/recursive. This arbitrariness of possible searches makes the decentralised time-space representation susceptible to mining, and as such to the production of new knowledge. With the persistent doubled articulation of such data-doubles, all communication acts in the European Union are continuously in circulation and within profiler reach — beginning with any act — even while the devices are presented solely as nodes of a static network laid down in a database. But the profilers' intent is to connect the dots — the data doubles of telecommunication acts. The system can connect any data with any data (simply because they have been generated) in a fully automated scanning mode through predefined algorithmic search mechanisms.

Lingua Franca in Data Space: Location Data

Location data of mobile phones might be the most important trace for the profiler, as location becomes the lingua franca in the surveillance and profiling community (Curry, 2004). This data is of specific value and shows most precisely the difference between metadata and the actual data used in data retention: ". . . as long as the phone is turned on, it serves as a passport into a monitored electromagnetic enclosure" (Andrejevic, 2007, p. 100).

---

11 D1 and d2 might be same database if both devices use communication services from one and the same provider. The bigger the provider, the more often this is the case. In a monopoly situation, every communication act is laid down in one single database infrastructure.

Location data is a necessity for the operation of the GSM standard that is solely used in the EU. Any mobile phone switched "on" produces locality-based data via the closest cells of the phone. This is needed to identify the phone in the provider's network and to offer the best connectivity available. However, retention of this data is *not* essential for the efficient functioning of the communication structure.[12]

To retain this data, to process it, and to make it accessible via searchable databases changes its status from ephemeral to temporally enduring. But additionally, retained location data remodels a four-dimensional world. Device locations are laid down as a space-time continuum of at least six months. Through algorithmic operations, these data can easily be visualised and brought into navigable and replicable form.

It is important to emphasise that location data derives not only through the explicit mobile phone connections; any landline call or communication service is locatable too, and the customer's data (such as address and billing information) are retained as well. The mobile phone's data offers profilers the luxury of the detailed history of a person's movement itself with remarkable precision as long as the phone is switched "on." This density of data results from the real world, from the material movements of bodies on earth. A one-to-one representation of the loci of any mobile device is the most accurate representation possible of people's movements and places (that can be achieved without invading the body itself). Therefore, participating in modern life in the 21st century is becoming increasingly dependent on a device that might also be called location tracker.

Tomorrow, Not Today, and Not in the Future

Once data retention is fully operational throughout the European Union, location data will be the ubiquitous source for the assessment of the population's movement.

By the principle of double (twofold) retention of all telecommunication acts, an operation on this data space is possible in any logical direction of the database's sets. Double retention in combination with location data may even serve to generate a more than four- dimensional world, as any topological construction is computable. The genesis of new technologically-oriented perspectives on the past (which will barely then qualify *as* the past) will then be cast in the looping circuitry of algorithmic operations.

But it is not only the past that is being re-mastered. With statistical methods, future predictions are also possible. As the base of these statistical calculations is a one-to-one representation and not a sample, the outcome of the prediction will be as accurate as statistics can possibly be. In other words: with the retention of location and traffic data, even the future of Europe's population is under (re-) construction. Groups and the movements of groups will be predictable at the pace of the latest CPU.

---

12 It is not a necessity for the member states to force the retention of this specific data. Location data in the strict EU directive's sense has to be retained only in combination with a successful or unsuccessful act of communication. Nonetheless, to retain all location data is a common practice in law enforcement.

Data Retention as an Element of an Assemblage

Seen from a broader scope, data retention fits well into the post-9/11 *war on terror* measurements. The shift toward an omniscient surveillance-state has generally often been compared to scenarios familiar from the prophetic novel *1984* by George Orwell (1949). But there are critical differences. The analogy had been underscored by the historical concept of the panopticon, introduced by Bentham (1785) and popularised by Foucault (1977). But where the panopticon draws its power from the fact that the surveilled never know if they are surveilled, and therefore internalise habits *as if* they were surveilled, the present situation, fostered by ongoing modalities such as data retention, should more accurately be referred to as *panspectron* or *surveillant assemblage*.

The term panspectron, as introduced by Sandra Braman (2006), refers to a state of things where no surveillance subject is specifically invoked in order to trigger an information collection process. Rather, information is collected about everything and everyone *all the time*. An individual subject appears only when a particular question needs to be answered, triggering data mining for particular information within the mass already gathered, in order to precisely answer that question. And while populations remain generally aware of the unmoved and intimidating presence of the panopticon, they tend to be unaware of the aggressive efficiency of specific modes of information collection. Data retention exactly fits into this conceptual frame, just as with Passenger Name Records and SWIFT financial data. These are sustained and augmented without any specified trigger, and therefore, potentially infinite.

The concept of the surveillant assemblage, as introduced by Kevin Haggerty and Richard Ericson (2000), refers to a multiplicity of heterogeneous objects, whose unity is solely functional. As an assemblage is always a "potentiality," this concept can be connected to the developing panspectron, which also resides in the background as a formidable and ambivalent latency.

Paradoxically, the weakness of the assemblage is also its power: "As it is multiple, unstable, and lacks discernible boundaries or responsible governmental departments, the surveillant assemblage cannot be dismantled by prohibiting a particularly unpalatable technology" (Haggerty & Ericson, 2000, p. 609).

The main directive of the assemblage is to transform the body into virtual bytes of information — data doubles. "And while such doubles ostensibly refer back to particular individuals, they transcend a purely representational idiom" (Haggerty & Ericson, 2000, p. 614). Discrimination and social sorting is amongst the socio-political consequences that subjects may experience as concrete back-references. Data retention is an example of a functional element of the assemblage that might refer back to subjects. It is unpredictable, and possibly holds radical consequences for the subject.

Conclusion: Communication is Doubled

The Royal Academy of Engineering speaks of necessary conditions for the application of surveillance technologies:

> Reciprocity between subject and controller is essential to ensure that data collection and surveillance technologies are used in a fair way. Reciprocity is the establishment of two-way communication and genuine dialogue, and is key to making surveillance acceptable to citizens. (Raeng, 2007, p. 8)

How can this be done? It is currently an open question as to whether a significant portion of European mobile phone users know about data retention at all or are sufficiently aware of its repercussions.

This is not only true of data retention, but also of most data collecting, processing, and mining practices. This significant lack of knowledge, understanding, and consciousness about new paradigms of technologically-enabled surveillance and related governmental practices and commercial business is comparable to a situation of betrayal: While the people communicating assume that their privacy is strongly protected, a permanent but virtual eavesdropping operation is in the making that might become all too real later.

The fact is that data retention of telecommunications data belongs to an assemblage of new emerging forms of control endemic to a networking society. Its power results from the wide acceptance and usage that electronic communications media have gained in the last years. Yet each link within this mega- or giga-weave of omnipotent cyber-connectivity is a single, tenuous, finite, vulnerable strand: the voice of the person at the other end of the line. To take part in this connected world is coupled with a number of drastic consequences that have been described above. Calling someone produces data far beyond the call. And while the call might be forgotten to have taken place within a couple of days by the people who have spoken to one another, the technical infrastructure implemented for data retention ensures for up to 24 months that it has happened, regardless of whether or not any human being remembers the call.

References

2006/24/EC. (European Union Directive 2006/24/EC of March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communication networks and amending Directive 2002/58/EC). *Official Journal of the European Union,* L 105, 56-63.

Andrejevic, M. (2007). *iSpy: Surveillance and Power in the Interactive Era.* Kansas: University of Kansas Press.

Bentham, J. (1995). *The Panopticon Writings.* London: Verso.

Bignami, F. (2007). Protecting Privacy against the Police in the European Union: The Data Retention Directive*. Duke Law School Science, Technology and Innovation Research Paper Series,* Research Paper No. 13.

Braman, S. (2006). Tactical Memory: the Politics of Openness in the Construction of Memory. *First Monday*, 11(7). Online-Publication.
URL: http://firstmonday.org/issues/issue11_7/braman/index.html

Curry, M.R. (2004). The Profiler's Question and the Treacherous Traveler: Narratives of Belonging in Commercial Aviation. *Surveillance & Society* 1(4): 475-499.

Deleuze, G. (1992). Postscript on the Societies of Control. *October* 59 (Winter): 3-7.

Foucault, M. (1977). *Discipline and Punish: The Birth of the Prison*. New York: Vintage.

Haggerty, K., & R. Ericson. (2000). The surveillant assemblage. *British Journal of Sociology,* (51)4: 605-622.

Krämer, S. (1992). Symbolische Maschinen, Computer und der Verlust des Ethischen im geistigen Tun. In W. Coy, W. *et al.,* (Eds.), *Sichtweisen der Informatik* (pp. 335-342). Braunschweig/Wiesbaden: Viehweg.

Latour, B. (2004). There is no Information, only Transformation. In Geert Lovink, *Uncanny Networks: Dialogues with the Virtual Intelligentsia* (pp. 154-160). Cambridge, MA.: MIT Press.

Orwell, G. (1949). *Nineteen Eighty-Four.* New York: Penguin.

Raeng (The Royal Academy of Engineering). (2007). *Dilemmas of Privacy and Surveillance. Challenges of Technological Change*. Online-Publication.
URL: http://www.raeng.org.uk/policy/reports/default.htm

Tholen, G.F. (1994). Platzverweis. Unmögliche Zwischenspiele zwischen Mensch und Maschine. In N. Bolz, F. Kittler, G.F. Tholen (Eds.), *Computer als Medium* (pp. 111-135). München: Fink.

Warner, J. (2005). The Right to Oblivion: Data Retention from Canada to Europe in Three Backward Steps. *University of Ottawa Law & Technology Journal.* 2: 77-104.

Winkler, H. (1997). *Docuverse - zur Medientheorie der Computer*. München: Boer.

Winkler, H. (2004). *Diskursökonomie. Versuch über die innere Ökonomie der Medien*. Frankfurt am Main: Suhrkamp.