Digital Enlightenment Yearbook 2013
M. Hildebrandt et al. (Eds.)
IOS Press, 2013

doi:10.3233/978-1-61499-295-0-289

Afterword

Kim CAMERON
Distinguished Engineer, Architect of Identity and Access, Microsoft

We who founded the Digital Enlightenment Forum wanted to create a deeper discussion of digital technology and its relation to society than was taking place around us. We had common interests but different professional backgrounds, which led to exchanges of ideas we thought valuable, not just for ourselves but, potentially, for others. We wanted especially to bring together four groups who would normally never spend much time together – scientists, technology innovators, legal experts and policy makers – and have them collaborate on understanding what was happening to our society.

We are living in a period of accelerating technological change. But there is relatively little understanding or discussion by people, otherwise keenly aware of the world around them, of how our new technologies actually work, of what they potentially effect, or of how they might impact the tenets of our society. Nor are technologies generally seen as things that can be managed: they are perceived as inevitabilities even though they are increasingly social and in fact are carefully managed to achieve conscious outcomes by all those who deploy them.

Given the absence of a conversation through which the public – or even influential elites – can begin to approach these issues, society's concept of itself risks being stuck in the past while reality changes underfoot. Sadly, this would mean democratic institutions would have no chance to shape their relation to these technologies. Just as ominous, the public may wake up one day in a very bad mood as changes that have become a *fait accompli* seep into their consciousness.

We saw omens of this even as this volume was being written, when a man as eminent as the President of the United States was swept up in a political whirlwind: revelations of digital surveillance shocked many by what they taught about contemporary uses of digital technology.

It is worth pausing to understand why Mr. Obama's handling of this political challenge stands as an omen rather than just a passing headline.

To put some limits on the discussion, let's remember that surveillance isn't new. I leave it to the many experts pro and con to lead what is absolutely an important debate about its necessity and legitimacy. Nor is digital surveillance unique to the U.S. The Guardian and Le Monde have asserted that many countries, including France, the United Kingdom and Germany, have digital surveillance programs similar to those operated in the U.S., while most countries that do not have such programs have constituencies within government that wish they did.

The situation challenging Mr. Obama was unique because what had been secret suddenly became public, concerned everyone using a telephone or the Internet, and was authoritative, consisting of actual very detailed government documents. The U.S. Administration judged that it was necessary to present the government's case for digital surveillance to the world.

As Mr. Obama began to do this, one of his statements stood out: 'This debate has gotten cloudy very quickly.'

Yes it did – and there are historical as well as political and ideological reasons why. Passage to the digital epoch involves many kinds of rupture, each creating inevitable confusion. In digital technology as in all else, things are not necessarily as they appear. We need to look beyond the surface of this new reality and arrive at an understanding of what explains it.

As an example of what I mean by this, let's turn to one the cloudier moments in Mr. Obama's own televised statement of why he supported the surveillance programs:

'(The 215 Program)… gets data from the service providers like a Verizon in bulk, and basically you have call pairs. You have my telephone number connecting with your telephone number. There are no names. There is no content in that database. All it is, is the number pairs, when those calls took place, how long they took place…'

Why cloudy? Because Mr. Obama, no matter how reputed as a law professor and talented as a politician, finds it convincing to first describe what is in the database and then say it has no content. This in spite of the fact that anyone who has ever used a spreadsheet will spot an oxymoron.

One might wonder if the koan of 'contentless content' were just a rhetorical device gone bad. But the expert communicator in Mr. Obama thinks he is connecting with his listeners. How is that possible?

Mr. Obama counts on a vocabulary in which artifacts that pre-exist the digital age are 'content' while those that are digital are not. He addresses his listeners from the pre-digital side of the chasm. Names are content, but phone numbers are not; phone calls are content, but the record of their time and length is not; the text of mail (even email because it is still… mail) is content; but the identifiers of the email recipients are not. Mr. Obama's discourse explicitly rejects making the transition to defining 'content' as including the artifacts that define digital life – including digital metadata.

Yet all those millions who have 'gone digital' in their daily personal lives know that their phone calls and the SMSs they send, combined with the times of day and durations of calls, reveal their social network. They know instinctively that the composition and usage of our social networks is as much 'content' as is a telephone call or video. As a result, the Internet seethed with descriptions of what was called Mr. Obama's duplicitousness: some polls showed disapproval reaching 61%; one found Americans under 30, who, according to conventional wisdom, care little about privacy, flipped against Obama by 17 points in a matter of days.

We need to consider the semiotics of Mr. Obama's challenge to see how the underlying issues should have been framed "now that we are digital". Social consensus around technology and society can no longer be fudged using the categories and conveniently naïve mental models of the pre-digital world: too many people in all age groups have digital insights, inchoate as they may be, and have begun to stumble upon new paradigms.

1. Identifiers and Databases

Let us return to the question of 'numbers' with no content. In a digital world, of course, everything is numbers, so numbers must be content. But some numbers are special in

that they are identifiers for things that exist in the physical world. The digital world can't exist without these identifiers.

For example, when we get a cell phone, our personal information is added to a directory and linked to our cell phone number – which is an identifier. Such directories are typically available to the public and are always available to law enforcement.

When a program like Program 215 collects cell phone 'numbers but no names', it is because a simple directory lookup or 'join' with the phone directory can be used to convert any number to a name. There is no advantage to storing a name when the name is already associated with each number and can be retrieved as needed. This reality is widely enough understood that Mr. Obama's 'numbers without names' approach didn't convince anybody.

For the same reason, a press release put out by the NSA a little later claiming good behavior for abstaining from collecting geolocation information rang hollow for many: an ever-increasing number of people intuitively understand that data bases are interconnected.

In reality, as is widely known, it was unnecessary to collect geolocation information because, like all law enforcement in the United States, the NSA already effectively has a directory that provides access to it. Geolocation information is universally collected and is readily available in the U.S. for lookup, at an hourly fee and without a warrant, using the phone identifiers that are being collected. A recent ACLU study indicates that most law enforcement agencies regularly rely on cell phone companies to track the physical location of their customers.

So to return to the question of why Mr. Obama's challenge was an omen, beyond the surveillance itself there were two serious problems with the way the U.S. Administration tried to explain it to the world. First, enough people have enough intuitions about the digital epoch to doubt the administration's explanations and render them counter-productive. Second, some of their explanations were intentionally framed to obscure the nature of digital reality, instead of helping people understand it. In this new confusing epoch, understanding is essential to society's wellbeing and trust.

This said, Mr. Obama made a fundamental breakthrough in political discourse when at one point he spoke to his television audience about the potentially identifying aspects of 'so-called metadata'. Who could have previously envisioned a head of state trading views on metadata with a popular talk-show host? In the wake of this conversation, metadata is becoming a household word that helps people understand the world they really live in.

Meanwhile, how many others take refuge in the sleight of hand that identifiers are 'just numbers' with no connection to a vast web of content. The same premise is used by many of the world's governments, and governments are not necessarily the worst offenders. Treating identifiers as 'just numbers' is all the rage among various big businesses that benefit from the currently veiled structure of the disparate databases which the numbers weave together.

2. Supercontent and Big Data

The identifying numbers that are used to connect people and things to the content located in a staggering maze of databases are the keys to the structure of the digital world.

To properly think about the digital age, we need both to recognize this identifying metadata as content, and, I propose, to promote it to the category of supercontent. Identifiers – numbers without names – are the supercontent that links all other content.

There are many different kinds of identifiers. More precisely, unique identifiers exist in different 'namespaces' or domains. For example, telephone numbers identify cell phones and land lines. Postal addresses identify places where people live and work. IP addresses identify computing devices connected to the Internet. Social Security Numbers identify individuals in a given country. Email addresses identify email recipients and originators. 'Login names' identify users of computers and applications. Customer numbers identify users of products and services. Social networking 'handles' identify members of social networks. Credit Card numbers identify credit accounts. Web addresses identify Internet services. All of these identify either a person or a device or resource used by a person or persons.

In the pre-digital world, namespaces were generally disconnected from each other. One of the key transformations we have seen in the digital epoch has been the stitching together of these identifiers across more and more domains.

When using a browser to place an order on the Internet one enters, at a minimum, one's name, address, phone numbers, email address, and credit card number. Meanwhile, one's computer transmits its current IP address, the website likely adds a cookie (identifier) in case the IP address changes, and more often than not web beacons associate the transaction with a set of other transactions made by the same user at other completely unrelated websites. In that single transaction, 8 namespaces are 'joined' together. If the identifying information is stored and can be consulted during subsequent transactions, then any transaction that reveals even one of these identifiers can do a lookup to determine the other 7. Further, if the new transaction contains additional identifiers, the new ones can be stitched in with those already assembled. In this sense, identifiers are 'contagious' supercontent.

As discussed in the chapters of this volume, our digital epoch is synonymous with the storage of the records of people's daily lives, their activities, transactions, profiles, appraisals, test results and evaluations in databases distributed throughout the world. The identifiers we have been discussing are what connect each of these database records to the person it describes.

Until now the records themselves have been distributed through space and controlled by different entities. They have been difficult and expensive to access and assemble because built on incompatible platforms, protected through disconnected security systems, and governed by data usage restrictions.

Against this background the World Economic Forum (WEF) has put forward a program that promotes 'Big Data' as an economic driver and a plan to 'rethink' (critics say 'eliminate') restrictions on the collection and use of personal data. The WEF report argues that 'The discovery and insights derived from linking previously disparate bits of data have become essential for innovation'. It urges allowing – in fact, promoting – the co-location and assembly of billions of records from innumerable databases, including the development of new interoperable mechanisms for doing so, to facilitate data mining and the search for patterns that might hasten advances in technology and science. In this plan, identifier supercontent (identity metadata) would be leveraged in order to 'unlock' the economic value of these interconnected fields of data.

From a technical point of view, it is not clear how many significant differences exist between the uses of identifiers and databases proposed by the World Economic Forum report and what is currently being done in various surveillance programs. This

may be why Mr. Obama, as part of his discussion of the issues, called for a national conversation on big data, arguing that 'the general problem of data, big datasets… is not going to be restricted to government entities.'

There is no doubt that the discussion of big data has also 'gotten cloudy very quickly'. Statements like 'Analytics have become the new engine of social value creation' are made without hard evidence, and need to be evaluated empirically. There has also been much overgeneralization – for example, using a useful outcome resulting from the connection of two specific datasets to argue against restrictions on connecting any data to any other.

The essential point is that any data containing identifier supercontent – an identifier that can be linked to a person directly or through contagion – is clearly personal data. The question to be answered by the Digital Enlightenment must be: under what conditions can personal data about people and their possessions be used for purposes other than the transactions they were collected to enable? In particular, is the premise that inference engines might detect associations that propel the human race forward – while bringing in the bacon – sufficient for unleashing what some see as a data free-for-all in which any data can be connected with any other?

This is also an area where philosophers of science could contribute to the Enlightenment discussion. The scientific method, which has led to virtually all scientific breakthroughs until now, begins with the elaboration of theories based on models that can then be expressed as hypotheses which are 'disprovable' if the theory is not correct. The hypotheses to be tested thus select the data that is to be generated or collated and analyzed for a specific purpose. In this paradigm, use of data can be subject to rational evaluation of risks and benefits. However the premise of the WEF report is that innovation is now no longer possible except through analytics, meaning the use of inference engines that graze the data collected (potentially by every device in the world) for associations and statistical patterns leading to huge new breakthroughs and value creation and eliminating the need for *a priori* hypothesis. The contention is that the scientific method as we have known it has been superseded – so control over data should be abandoned as we get on with the grazing epoch. However, since no empirical evidence is provided for this assertion, it seems worthy of more discussion before rushing to comply.

3. Data Reciprocity

One related area no one seems to be discussing is the idea of "Data Reciprocity". Rather than eliminating restrictions on the collection of data, Data Reciprocity would encourage restrictions on collection, combined with a requirement that any data collected from an individual that contains identifying supercontent must be made available to that individual through an Application Programming Interface (API). Such APIs would attract developers to develop applications for end users that would 'join' their own information records for their own use. For example, if someone opens an account at a retailer and places orders, it should, given Data Reciprocity, be possible for that person to retrieve all the information he or she has generated with the retailer through APIs.

The development of Data Reciprocity APIs would be many orders of magnitude less onerous than what is proposed by the WEF report calling for all information anywhere to be sucked into Big Data systems.

Data Reciprocity should include a requirement that governments and private enterprises collecting email addresses register the location of their Data Reciprocity API in a registry associated with each person's email address, and only accessible to the individual identified and the software services he or she designates. This would make it possible to develop systems to assess compliance with Data Reciprocity requirements.

A rich area for research and discussion within the Digital Enlightenment community could then be whether and under what circumstances Data Reciprocity should apply to information derived from personally identifiable information and made available to third parties. For example, if categorization of an individual's economic status is done using personally identifiable information subject to Data Reciprocity, should it be a requirement that the derived economic status categorization also be made available to the subject concerned via an API? In any case, establishing the principle of Data Reciprocity is one of the most important enablers of "Life Management Platforms". It also serves to attach a cost to the collection of personally identifiable information without being punitive in any way.

The articles the Digital Enlightenment Forum has brought together in this volume provide much insight that can be applied to these and other related issues. But while we continue to deepen our own understanding of the world around us we must also address the question of how we connect with the increasing number of people who are spontaneously coming to understand more about how digital society works. Digital enlightenment can mean nothing less than a society fully conscious of its digital world and the social issues that characterize it. Standing for a democratic society that is fully digital yet consistent with our values, Digital Enlightenment is more than a name. It is more than a goal. Digital Enlightenment is a necessity.