Künstl Intell (2012) 26:279–283
DOI 10.1007/s13218-012-0193-4

INTERVIEW

Interview with Josep Domingo-Ferrer and Chris Clifton on Privacy in Spatiotemporal Data Analysis

Christine Körner

Published online: 28 April 2012

Josep Domingo-Ferrer is a Full Professor of Computer Science and an ICREA-Acadèmia Researcher at Universitat Rovira i Virgili, Tarragona, Catalonia, where he holds the UNESCO Chair in Data Privacy. He received his M. Sc. and Ph.D. in Computer Science from the Autonomous University of Barcelona in 1988 and 1991 (Outstanding Graduation Award). He also holds an M. Sc. in Mathematics. His research interests are in data privacy, data security, statistical disclosure control and cryptographic protocols, with a focus on the conciliation of privacy, security and functionality. He has authored 5 patents and over 270 publications. He has led major national and international research projects on privacy. He is a co-founder and a co-editor-in-chief of the Transactions on Data Privacy journal. He is an IEEE Fellow.

C. Körner (✉)
Knowledge Discovery Organization, Fraunhofer IAIS,
Schloss Birlinghoven, 53757 Sankt Augustin, Germany
e-mail: christine.koerner@iais.fraunhofer.de

Chris Clifton works on data privacy, particularly with respect to analysis of private data. This includes privacy-preserving data mining, data de-identification and anonymization, and limits on identifying individuals from data mining models. He also works more broadly in data mining, including data mining of text and data mining techniques applied to interoperation of heterogeneous information sources. He also works on database support for widely distributed and autonomously controlled information, particularly issues related to data privacy. Prior to joining Purdue in 2001, Chris Clifton was a principal scientist in the Information Technology Division at the MITRE Corporation. Before joining MITRE in 1995, he was an Assistant Professor of Computer Science at Northwestern University. He has a Ph.D. (1991) and M.A. (1988) from Princeton University, and Bachelor's and Master's degrees (1986) from the Massachusetts Institute of Technology.

KI: Given that hundreds of millions of users are apparently ready to share very detailed private information on platforms such as Facebook or Twitter, isn't the concept of privacy an obsolete notion today?

Josep Domingo-Ferrer: In the information society, privacy can no longer be understood as seclusion or isolation. Hardly any active participation in the information society is possible if one wants to stay isolated. The current notion of privacy is rather one of control on what information on ourselves is seen by whom, and when, where and for how long it is seen. This requirement of dissemination control is far more challenging from the technical point of view than the old idea of seclusion.

Chris Clifton: There are two issues with calling privacy obsolete. First, the information users provide on such services is what they choose to provide. However, a lot of information, including mobility data, is collected with little choice on the part of the individual. Second, users often don't realize how much they are giving up. While this is improving (witness Google's move from a plethora of privacy policies to a single 2400 word policy—not counting special policies for Chrome, Books, and Wallet), it is still debatable if users are giving *informed* consent. Informed consent means that if you provide data, on each use of the data you would have to be told what the data specifically would be used for and you would have to be asked to grant consent for any uses that are not explicitly allowed under the privacy parts. Informed consent is a big issue, and I think the European Union has a better handle on that.

KI: Do you see differences between societies on different continents in how they approach privacy? How will the relevance of privacy develop over the next 10 years in Europe respectively America?

Josep Domingo-Ferrer: Indeed, even if privacy has been recognized as a basic human right in Article 12 of the Declaration of Human Rights (1948), its perceived importance strongly depends on the culture. In Europe, perhaps due to many countries being small, privacy is regarded as a key value. My impression is that, in the United States, privacy is more viewed as a commodity, that is, an added value for consumers that must be made profitable for companies to invest in it. Outside the "old developed countries" (i.e. outside Europe, North America, Australia, New Zealand and maybe Japan), individual rights in general and privacy in particular are deemed less relevant. Hopefully, privacy awareness and demands will increase with democracy in those regions.

Regarding the future in Europe, according to recent news, the European Commission is finalizing the first significant update in privacy legislation since 1995. The new directive will strengthen the EU powers to fight data protection breaches. Approval of this update is likely to take two more years, with another two needed for deployment. In 10 years, there is a big opportunity for Europe to consolidate its leadership in the creation of privacy-aware ITC technologies.

Chris Clifton: I see some differences in societal views, for example, the U.S. shows more concern about government intrusion than many countries. I think the reason for this is largely historical. If we go back to the founding of the country and the start of the U.S. legal system as being an independent legal system, people lived in an environment where there were serious abuses by the colonial governors and so the U.S. constitution contains passages protecting against government intrusion.

In the U.S. we tend to react to what happens. I think that we will see growing concern as word of uses, or misuses, of private data spread. For example, mobile data (in this case, toll pass records) have been used in a civil court case in the U.S.—and people are becoming more aware that such data is being collected and can be used against them. This must be balanced against the convenience in new services based on collection and use of private data. Another, extreme, example is the Arab Spring, which let people realize that there is both a lot of power in new media and also a lot of risk. I think we will continue to see new uses of data that pose great privacy risks becoming the "hot new thing", well before people realize the privacy risks. This reactive approach is opposed to the proactive approach to privacy that the European community has taken.

KI: How do you see the legal situation around privacy in different regions of the world? Are there marked differences in laws and regulation?

Josep Domingo-Ferrer: In the United States, privacy is legally protected only to the extent that the lack of privacy jeopardizes the constitutional right to freedom of speech. It is also protected in the field of official statistics, in order to reduce the non-response rates. In the big Asian countries, like China, India, etc., there is less emphasis on the individual and more on the group, on the society. Hence, privacy is not regarded as a central value. There is a risk that emerging countries overlook privacy regulations the same way they have sometimes overlooked environmental regulations. In Europe, privacy enjoys strong and hopefully increasing legal protection both at the EU and the national levels.

Chris Clifton: I see greater differences in legal approaches than in societal norms. In the United States we have long-standing laws protecting privacy against governmental intrusion. However, it is a matter of open debate if there is a general "right to privacy" under U.S. law. Instead, most states have sector-specific laws—one law for email, another for telephones, for educational records, health data, etc. These laws are similar between states, but not identical. There are cultural differences in different parts of the

U.S., and some areas adopt privacy rules faster than others. Eventually, some of these laws become unified; but as the states tend to move faster than the federal government, we will always have a patchwork of privacy laws.

One bright spot in U.S. law, at least for researchers, is the well-known Health Insurance Portability and Accountability Act (HIPAA). To my knowledge, this is one of the few laws that provides a good way to quantify "individual identifiability"—the safe harbor rules specify exactly how data can be generalized to meet the legal requirements. This gives a researcher a starting point to analyze what privacy means in terms of risk.

KI: Do you think privacy regulations constrain and slow down innovation and growth of new services?

Josep Domingo-Ferrer: I will answer with another question. Do you think that environmental regulations constrain and slow down industrial innovation? Privacy preservation is in many respects parallel to environment preservation. In fact, privacy preservation is about preserving our personal information ecosystem. For the information society to stay human, it needs to be privacy-aware. In my view, privacy-aware innovation is more challenging and, hence, more innovative, just like low-emission car engines are technically more advanced than the old powerful and wasteful engines.

Chris Clifton: Yes and no. Yes, privacy regulations slow down innovation and growth. I think they slow down growth more than they slow down innovation. People come up with new services without realizing privacy implications. They try them out and if there is a problem it gets killed. So I think it has much less impact on innovation because a lot of innovators do not think about privacy issues. On the other hand privacy regulations can also drive new products and services—such as breach disclosure laws in the U.S. leading to new identity protection and credit protection services. We have not seen a lot of research put into practice—not as much as I would like to see—but hopefully that will come.

KI: How do privacy requirements affect the competitive position of individual companies? Are there advantages to be gained from not worrying about privacy, or are there companies which actually have a competitive advantage because they take privacy seriously? Can you give best-practice examples of companies?

Josep Domingo-Ferrer: Ignoring privacy certainly gives immediate gains resulting from personal information exploitation (profiling, market segmentation, advertising). I claim that much of that information exploitation can still be done in a privacy-preserving way: privacy-preserving data mining (PPDM) and statistical disclosure control are precisely about this. Of course, it is a bit more complicated to use PPDM than plain data mining, but the overhead can decrease with automation. On the other hand, if we adopt the American view of privacy as a commodity, IT companies could gain a competitive edge if they marketed their products and services as "privacy-aware". A "privacy seal" could be bestowed to companies, products and services by independent accreditation agencies, in a similar way environmental or quality accreditations are now granted and proudly displayed. If people are ready to pay a bit more for a hybrid car, wouldn't they pay a bit more for a privacy-certified product or service?

Lacking the above privacy certifications for products, best privacy practices are to be found in administrations rather than in companies. E.g. most national statistical offices in Western countries are subject to strict privacy regulations and nonetheless they do their job efficiently while guaranteeing privacy to citizens.

Chris Clifton: Citibank is my favorite example of a company using privacy to give competitive advantage. They use privacy concerns to drive new services (most of which truly serve as fraud prevention tools). Some of these are questionable from a privacy point of view, e.g. putting your photo on your credit card. But others, such as single-use credit card numbers, can provide real privacy advantages. In this case, I can go to a Citibank website, log in and have it generate a credit card number, which is tied to my account. When I make a purchase with a merchant, the company gets a credit card number, and I never use it again. So they are unable to use that credit card number to track my history of purchases. This is providing privacy and it is giving me a real advantage. Additionally, it gives Citibank fraud prevention because if that number is disclosed and someone tries to reuse it, it does not work.

So I think there are real innovations in practice that provide privacy advantages. However, consumers are not willing to pay enough for privacy to drive innovation purely out of privacy reasons. Thus, companies use opportunities where they can sell privacy technology not because it is providing privacy but because it is providing some other benefit that people are willing to pay for.

KI: Is spatial and mobility information particularly sensitive when it comes to privacy? Which emerging applications and services do you consider especially critical w.r.t. the privacy of personal spatial and mobility information?

Josep Domingo-Ferrer: Indeed, it is. Your whereabouts tell a lot about your social relations and your lifestyle. You cannot prevent automated collection of your trajectory by cell phone operators, for example. But these should not release any mobility data in a way that allows re-identifying individual trajectories. Service providers other than cell phone operators, like location-based service providers, do not need to be told your exact position. There is a large body of literature on how to provide mobility services in a privacy-preserving way.

Chris Clifton: There are some critical issues. One of the strongest U.S. privacy laws deals with homeless shelters—primarily to protect battered spouses who can be in real physical danger if discovered. Mobility data is even more critical in this respect. If an abusive spouse can figure out exactly where you are from your cell phone records, you are going to be in trouble. There was also a court case in the U.S. where they used someone's toll pass records to show that they had driven to some other place than they said they were going to. That is clearly an example where mobility data has been used in a way that the person moving would not have wanted. Whether that is good or bad is a matter of debate. But it certainly shows that there are privacy risks with respect to spatial and mobility information, and that these are real.

KI: Do we need special legal regulations for spatial and mobility data? Do some countries already have them?

Josep Domingo-Ferrer: I am not aware of legal provisions specific to mobility data, but I am not a legal expert. I think mobility data are a form of personal data and they can be protected by the same legislation that protects personal data. The difference between protecting mobility data and ordinary personal data regards technology rather than law. National statistical offices are among the most privacy-conscious organizations, but most of them are not yet used to collecting mobility data and releasing them in anonymized form. And the private sector, whose job is not releasing data, is even less interested in anonymization of mobility data. What is certainly essential is that data protection agencies take serious action against any organization or company releasing re-identifiable mobility data: they must make it clear that a trajectory is as private as a medical record.

Chris Clifton: I believe there are some such laws in place. The U.S. has a patchwork of individual state laws, and some may cover mobility data. However, to my knowledge there is no U.S. federal law that covers mobility data in general. I expect that for the near future, at least, U.S. "law" on mobility data will either be at the individual state level, or based on attempts to apply existing law such as the Electronic Communications Privacy Act (which is really set up to protect against intercepting message content). As people then see that existing law does not do a good job on mobility data and we have a few high profile examples of misuse, we will start to see changes in the legal system.

KI: What are the main technical challenges for a safe handling of sensitive spatial and mobility information, and are there already approaches that are promising concerning the privacy-preserving exploitation of spatial and mobility data?

Josep Domingo-Ferrer: A first challenge is that the anonymization methods used for conventional microdata are in general not suitable for mobility data. Even privacy models, like k-anonymity, must be redefined and adapted for mobility data. The reason is that, in trajectory data, any point and/or time can be regarded as a quasi-identifier, that is, an information that can be used to link that trajectory with the identity of someone who is known to have been at that point at that time. Hence, a second challenge is to develop new models that precisely capture trajectory privacy and new methods that satisfy those privacy models while providing enough utility in the anonymized data. There are indeed some methods in the literature that are promising and achieve some form of trajectory k-anonymity via microaggregation and/or synthetic trajectory generation. Improving those methods in view of increasing the utility of the resulting anonymized trajectories (less data suppression, for example, while better preserving trajectory shape and compatibility with the underlying road network) is among the open research issues.

Chris Clifton: I think the biggest issue is the inference problem—given some mobility data, and some external information, much more may be inferred. Trying to control this through laws will be difficult—we can't outlaw intelligence (although in the U.S., there have been attempts such as the Digital Millennium Copyright Act). I think there has been some solid work done in anonymization, meaning that you become indistinguishable from a crowd. Recently also the idea of differential privacy has emerged, which deals quite well with the inference problem. The idea behind differential privacy is that when querying a database noise is added to the result, which hides the contribution of anyone individual. One of the big problems with differential privacy is that quantifying how much noise has to be added is based on a parameter, and the parameter does not relate to the risk of disclosure of individual data in any recognizable way. At this point differential privacy is a nice idea and has a lot of power with respect to mobility data but it is not ready for practice yet. One of the biggest questions now is understanding what the risk is and when the technologies are sufficient to adequately protect privacy.

KI: Do you think the general public awareness towards the necessity of privacy should be strengthened? Which steps are already undertaken in Europe respectively America or should be undertaken?

Josep Domingo-Ferrer: That's indeed fundamental. The citizen should be aware of the dangers of being careless about privacy. The idea "I don't do anything I need to hide" is flawed. Maybe you don't need to hide now your parties with your friends, but you might be embarrassed at them in 10 or 20 years, when, for example, you talk your children into going out less often. Or you might be embarrassed by having owned a 4WD vehicle if you later become a green

party leader. In fact, if you later become someone uncomfortable to powerful corporations, organizations or countries, your digital traces and profiles are likely to be used against you. Everyone has the right to be forgotten and the right to start again in life, and these are aspects of privacy preservation. Alternatively, if these arguments do not convince you, just think: if companies reward me with fidelity cards in order to profile me, that means my privacy is worth something.

As I said above, the EU is updating its privacy regulations and, if data protection agencies actively enforce this new legal framework, the creation of privacy-aware technologies will undoubtedly be stimulated.

Our modest contribution from the UNESCO Chair in Data Privacy is to gather and publish in our web site http://unescoprivacychair.urv.cat any privacy news that we judge socially relevant.

Chris Clifton: We do have our privacy zealots in the United States, who try to educate people on risks. A few still remember historical abuses, such as the McCarthy hearings and blacklisting in the movie industry based on rumors of communist sympathy or activity—but few really believe this could happen today. However, we also have newer examples, such as companies scanning Facebook pages, and archives, when evaluating job applicants. Our (student-run) school newspaper runs periodic editorials telling students of the dangers—consider your Facebook page part of your job application. So I think people are learning.

The key to educating the public is to be able to point to real abuses with real harm. People would rather have the benefit and not worry about the risk. So showing the potential for loss of privacy isn't convincing—we have to wait for something bad to happen before people really listen. But when it does, being ready with the right technology is important—to avoid public backlash and laws that throw out the good uses along with the abuses.

KI: How do you personally protect your privacy? What do you recommend?

Josep Domingo-Ferrer: I don't use social networks, just e-mail. I don't use Gmail or Yahoo e-mail accounts, although I cannot avoid corresponding with people who do. I don't share many pictures in the cloud, but if I occasionally do, I never assign meaningful names, labels or metadata to the images. I regularly eliminate cookies and other tracking data from my browser.

Also, you will only find professional information in my personal website, albeit detailed one. I cannot avoid using search engines, although I bookmark my preferred sites in order to access them directly without looking them up in Google.

I don't express opinions in microblogging services like Twitter. Whenever I have expressed personal or political opinions in press articles, I have done so after pondering whether my privacy disclosure was outweighed by the foreseeable benefits of disseminating my stance.

In general, my advice is: make rational decisions when disclosing information and keep in mind that what you post on the Internet will be seen by anyone (including your potential enemies, rivals or competitors) and probably forever (unless we succeed in deploying digital forgetting mechanisms for good).

Chris Clifton: At this point, the best way is to not generate information I don't want released. For example, I carry a cell phone only when I think I might have use for it, and only power it on when I am making a call or expect to receive one. I change identifiers (e.g., the MAC address on my computer) periodically. And I do request (such as through refusing cookies) that places don't track me.

I also avoid services such as Facebook whose privacy policies I'm uncomfortable with. This is more of an educational ploy—"I'm not on Facebook" is a good conversation-starter into the reasons to be careful with such services, than any feeling that I would use them in a way that would be a privacy issue.

More importantly, if I were to do something I didn't want tracked, I'm aware of a variety of ways to avoid it. For example, in the U.S. it is possible to buy debit cards with cash, and without providing identification. Privacy isn't dead, it just requires knowledge, forethought, and sometimes, inconvenience and expense.

KI: Thank you very much for the interview!

Both interviews were held independently of each other.