Human-Centred Identity – From Rhetoric to Reality

Adrian Rahaman, M. Angela Sasse
Department of Computer Science
University College London, UK

Abstract

This paper presents a proposal for human-centred identity management. Even though the term 'human-centred identity' has been widely used in the past few years, the solutions either describe a technical system for managing identity, or describe an identity management solution that meets a particular administrative need. Our proposal, however, presents a set of propertis that have to be considered, and the choices have to be made for each property must satisfy the needs of both the individual and the organization that owns the identity management system. The properties were identified as a result of reviewing a range of national identity systems, and the problems that arise from them.

1 Introduction: Identity Systems Today

Identity is a construct that underlies the mechanisms which enable or prevent an individual from performing certain actions in a social environment. Either explicitly or implicitly, organizations have always sought to close 'the gap of uncertainty' between an individual's claimed identity, and their 'true' identity one. This is in no small part due to the increasing disembodiment of transactional processes – interactions that previously used to be conducted face-to-face, and using physical documents as evidence, are now mediated through information and communication technology [1, 2]. There has been a flurry of research activity in the area of identity and identity management, and many large-scale systems have been proposed, designed and implemented.

In the private sector, the increase of identity and information collection is driven by the wish to personalize services. For recommender and social networking systems, the aggregation of various types of information about individuals is essential. The public sector is using similar approaches to realize the ambition of the vision for "citizen-centric services", as well as datasharing to reduce costs and detect fraud. In the UK, security challenges (terrorism, crime, fraud, etc.) have led to increased monitoring of citizens and their activities, to the point of what critics describe as a Surveillance Society [3].

Individuals have accepted some of those developments, but voice their disagreement in other cases: e.g. Facebook users when profile updates were broadcast [4], or the public concerns about the introduction of national identity systems [5-7]. In some cases, there has been simple lack of adoption of certain technologies, e.g. the case of the Austrian eID [8]. The problem is that the systems have been based on what is technically feasible, and ignoring human needs and concerns [9]. This result in a lack of understanding as to how people actually view, assess and interact with identity systems. In this paper, we aim to identify how the needs and concerns can be addressed during the design stage.

1. 1 Identity and Privacy

The use of identity is a dialectic process: it involves the transfer of information from one party to another in order to progress through the various steps of identification, authentication and authorization. This immediately raises issues of control over, and safety of the identity information. Who should have access to what, when and why? These are the main issues that privacy research seeks to understand and solve.

Privacy studies wrestle with a multidimensional concept [10-12]. There is no single agreed definition of privacy. In the area of identity management systems, the focus is on informational privacy constraint [13] as identity is typically defined as a set of information/attributes, about an individual that sufficiently differentiates the individual from a set of other individuals.

Previous research on informational privacy dimensions has helped to build a legal body of rules to protect the subjects of such schemes; the cornerstone of which are the Data Protection Act and the Fair Information Usage principles. The rise in the level of privacy concerns have also resulted in the development of Privacy Enhancing Technologies (PETs), which aim to protect an individual's privacy in digital interactions by limiting or encrypting identifying information. However, while these solutions can and do help to address the concerns brought about by identity systems, they are not without problems.

One of the drawbacks with the privacy approach is that individuals make their decisions based on the *perceived* rather than *actual* level of privacy provided. Additionally, it has been shown that the line drawn between public and private is dynamic – it changes depending on the information and context of use. Individuals have claimed that certain information is off-limits, but disclose it when a trivial benefit is offered [14,15,13].

The various legal definitions of privacy are constructed at a higher level of abstraction than identity management systems. Attempts to apply legal constraints to system design usually reduces privacy assessment to a set of checklists (e.g. Privacy Impact Assessment checklists), rather than understanding the impact of the system on the individual. Privacy-enchancing technologies (PETs) are well-intentioned, but are yet again a technology-centered paradigm that means that the system designers and owners can use to absolve their responsibility with understanding what the impact on the lived experience of individuals will be.

1.2 Identity and Trust

Trust is required in situations of risk uncertainty. In an identity scheme, trust helps an individual to make decisions about disclosing information that might result in undesired usage (e.g. information abuse, identity theft). There has been much work in developing models of trust that aim to predict user decisions to take action in uncertain situations.

In attempting to create a trust model for national identity systems, [16] explored the intention to adopt such systems as the development of trust through several stages of the subject's interpretation of the situation. Another effort in the area comes from [17] in measuring the level of citizen trust towards authorities in the implementation of a European Union wide identity scheme.

While these approaches are useful in understanding which general areas can be improved to generate trust, it fails to account for the structure of an identity system itself. This approach provides very little linkage back to the actual identity system, and hence offers implementers little guidance on how the actual design of a system might influence behavior or perceptions.

2. Human-centred Identity – what is it?

Neither privacy or trust research can provide an answer to this question. Analyzing schemes from a privacy or trust perspective abstracts the identity system from the specific consequences that is has on individuals lives, how they interact with the system and the various 'coping' strategies that might be adopted.

Practitioners and researchers require a way of predicting the lived experience that results from participating in an identity ecosystem. Thus, they need tool or method that allows a system owner or developer to assess how the design of the identity systems might influence user interaction, perception and reaction.

2.1 Methodology

A tool aiming to assess the impact of an identity system design should be expressed as a set of 'configuration' properties into which any such system can be decomposed. We identified these properties through a review of past National Identity Systems. The scope of work is limited to National Identity Systems in the Western world, since information is readily available, and these countries have been leading the adoption electronic identity systems [18]. Each system was treated as a unique case study.

Thematic Coding [19, 20] was used to identify similarities across the narratives of past and present national-scale identity schemes. The analysis revealed that a system configuration can be broken down into two different attribute sets, i.e. the structural properties and the metrical properties. The individual properties from these respective sets 'measure' the amount of relevant affordances that the system can provide for each property.

2.2 Structural properties

The structure of an identity system refers to the manner in which an identity ecosystem can be constructed. These properties seek to capture the flow of information inside the web of identity that is established. Therefore, the structure of an identity scheme will define how the interaction between individual and society is shaped by the identification system. We now present and explain the structural properties that emerged from the analysis.

2.2.1 Control Points

One of the main structural properties of any identity system can be expressed in terms of the number of control points that is built into the overall scheme. Control points are defined as *the situations in which an individual's identity is required in order to proceed with a particular function*. When an identity ecosystem contains a large number of control points - where an individual's identity is required to move from one state to another - the identity is exposed frequently to the relying party. Inversely, a low level of control points implies that an individual's identity is not requested frequently.

2.2.2 Subject Involvement

This property captures the role played by the individual whenever his/her respective identity is consumed by another party across all possible control points – whether it is active or passive. A system with a high level of involvement gives individuals an active role in the presentation of their identity, i.e. an individual will need to be present when their identity is used. On the other end of the spectrum, individuals can be completely passive members of an identity scheme. Systems that make use of a centralized database to store information are prime candidates for low involvement. The records stored on the database can be accessed by the organization without the individual being present, and unaware that the identity is being accessed.

2.2.3 Discreetness

An individual enrolls into an identity system to gain access to certain resources - this involves the presentation and use of subject identities at various control points. This has the implication that an individual's identity, and the information attached to it, will be exposed to a consuming party. This process harbours the risk of identity "leakage" to non-consuming or non-reliant parties. Unnecessary disclosure of information at the various control points can be expressed as the level of discreetness of the identity system; as such, it refers to the level of control that individuals have in presenting the identity to the rest of society. A system with a low level of discreetness constantly "leaks" identity information to third parties that have no right or no permission to the identity. Identity systems that preserve the integrity of the identity from other parties offer high levels of discretion.

2.2.4 Population Participation

Finally, the level of population participation represents another structural property of an identification system. This property refers to the number of individuals that are enrolled and interact with the system, in relation to the size of the total population that participates and acts in the context of which the identity system operates. A system with a low level of population participation would be one that is highly targeted, where the number of subjects that are enrolled into the system consists of a small fraction of the entire population in that context. On the other hand, a system that by default has everyone in the population enroll has a high level of population participation.

2.3 Metrical Properties

The metric of an identity system refers to the various techniques, methods and technologies that are used to capture and present an individual's identity. The metrical properties defined here attempts to capture how individuals interact with, and are affected by, various affordances that the underlying identifying technologies of an identification system can offer. These attributes can serve to influence the behavior and perceptions of individuals that encounter identity systems.

2.3. 1 Comprehension

Firstly, there is the matter of individuals' comprehension towards the various metrical technologies and techniques used for identification. This property is expressed in terms of how well an individual's understanding of the identifying technologies is aligned with reality. A system that has low levels of comprehension is one where individuals do not understand how the metrics are used to identify them. If individuals have some idea of how the mechanism works, but they are not aware of the entire process, this still results in low levels of comprehension. Low levels of comprehension occur when individuals are unable to point out, explain or rectify any problems that might occur during the identification process. On the other hand, systems with high levels of comprehension are those in which an individual has a good mental representations of the process in which the identity metrics are used.

2.3.2 Expert Analysis

Another metrical property - related to subject comprehension - is that of expert analysis. This property refers to the amount of human activity engaged in making use of the information collected for identification purposes. Completely manual systems would equate to a high level of expert analysis as it requires 'experts' to handle the identifying metric at various stages throughout the lifecycle of the identity. As such, systems with high level of expert analysis typically result in highly subjective systems where the identity is dependent on the interpretation of information by human users. Automated systems serve to decrease the amount of expert analysis involved, providing systems with an objective approach to processing identity.

2.3.4 Information Accuracy

Information accuracy is a property of the metric that defines how reliable the system is in producing correct matches in the process of identification. Identity systems that offer high reliability in providing correct matches are said to be provide high level of information accuracy. However, this accuracy must not be based solely on the theoretical possibilities that have been touted for any particular identification metric. Accurate "measurement" of information accuracy will need to take into account the implementation specific details that can affect the theoretical figures that have been put forth. The inconsistencies and practical limitations of the real world will need to be reflected in the information accuracy property of the system.

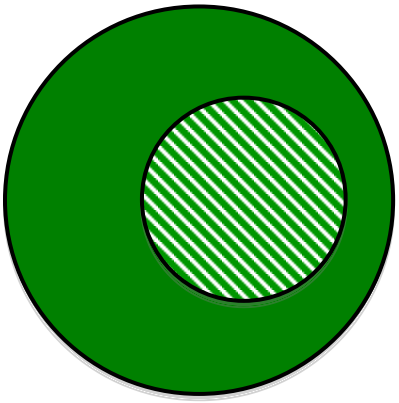
2.3.5 Identity Stability

The chosen metric for an identification system will also have an impact on the stability of the registered identity. Stability refers to the rate with which an individual's information stored in an identification system changes over time. A system has a low level of stability if the information associated with the identity has the potential to fluctuate greatly over short periods. Identity systems that make a large use of biographical information typically have low levels of stability as the information can potentially change at any given time (e.g. address, profession, even name). Conversely, purely biometrical systems can provide identity solutions with high levels of stability (depending on the biometric; facial recognition for example would not provide high levels of stability) as the metric is believed to remain constant over the lifetime of an individual.

2.3.6 Subject Coupling

Identification systems do not only vary in terms of the stability of the information collected, but also in terms of the amount of information that is collected and used for a particular purpose. This property of the system is known as subject coupling, i.e. the level of representativeness between the captured identity and the relevant 'partial identity' [21] of the subject in relation to the purpose and context.

A tight coupling suggests that the captured identity metrics faithfully represents a person's partial identity at the various control points that it is applied. On the other hand, a system that collects and reveals too much or too little information about an individual is said to have a low subject coupling, since the identity that is captured and presented does not accurately represent the 'complete' individual in that situation. While this property may seem like an easy aspect to establish, ensuring that subject coupling is accurately assessed depends on more subtle nuances about the information around the identity and the context.

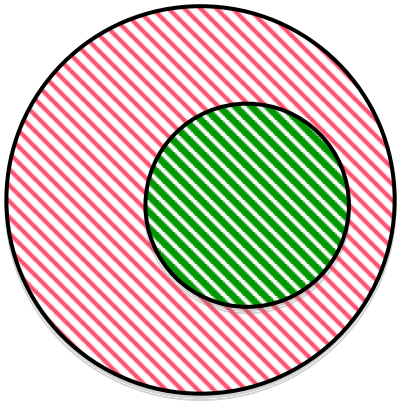


Low subject coupling due to a lack of information.

The identity consumer cannot come to an informed decision based on the information available.

Low subject coupling due to the availability of too much information.

The identity consumer runs the risk of passing judgement based on information unrelated to the context.



While a lack of information to represent an individual means that there is a low subject coupling, the inverse is not always true. As per the definition of this property, subject coupling occurs when the identity created does not represent the person in the context. This includes cases of under-representation as stated earlier but also that of over-representation. When 'too' much information is known about an individual the consumer of that identity might then judge the individual based on the information that is irrelevant.

2.3.7 Information Polymorphism

Depending on the chosen metric, an individual's identity may be more or less prone to being used for purposes that deviate from the original intention for which it was collected. The likelyhood that the identity may be used for some completely different purpose stems from the various meanings that can be attributed to or extracted from the information held about individuals. This is captured by the term *information polymorphism*. Systems with a high level of information polymorphism are those in which an individual's identity and information can be easily taken out of context of the original scheme, and applied to other systems that have completely unrelated purposes. Systems with a high level of information polymorphism lead to situations of function-creep of the identity. Alternatively, a low level of information polymorphism means that an individual's identity is safe from being exploited for other functions.

2.4 Combining Properties

Looking at the various properties individually can help researchers and practioners to understanding how and why the individuals might react to the introduction or alteration of an identity system. A system with a high number of control points might be perceived as – well: too controlling – and thus meet with resistance. A system with a low level of discreetness will be perceived as a violation of privacy, because the identity may be broadcast to parties that have no right to such information. Systems that need to be up-to-date but made use of a metric that has a low level of identity stability may be seen as a burden upon individuals, who continuously have to report when information changes.

We have not yet developed a complete map of interactions between properties, but feel that the explanatory power lies in the combination of the properties of interest, and observing the potential effects. For example, if one were to take a system with a *low population participation*, coupled with a *high subject involvement* and a *high number of control points*; this can lead to a scenario where a subject might be forced to abandon his/her 'identity' and construct a new one (if possible).

The identity system is a *highly targeted* one, indicating that certain criterion needs to be met for inclusion into the system. The majority of the population acting in that particular context is able to bypass the system; if individuals play an active role at a *large number of control points*, some individuals might come to the conclusion that the burden of the system is unbearable. As such, in cases where it is possible to do so (e.g. identification systems based on religion), it can be expected that a number of individuals might avoid the identity system altogether.

A final example: in an identification system with *low subject coupling*, *low understanding*, *low information accuracy* and a *high expert involvement* can lead to scenarios where subjects lose all 'power' leading to claims and actions made on incorrect interpretations of the identity. The low subject coupling means that the captured identity does not fully represent the individual in the context of the identification system. The high expert involvement and low information accuracy further degrade the quality of the identity and the process of identification. As a result, false accusations may be made against the individual, based on flawed conclusions drawn from the identity.

However, the fact that there is very little understanding regarding the system by non-experts, the possibility of successfully disproving any claims is reduced significantly. Therefore, a system with such a configuration will likely result in situations where subjects lose all ability to resist claims based on the identity. This results on incorrect actions taken against a subject that and can cause permanent harm.

3 Applying Properties to Real-World Scenarios

The system properties above were developed through an investigation of National Identity Systems. To illustrate the applicability of the properties to different contexts, the properties will be used to investigate identity and information systems that have been implemented in completely different environments. In the follwing, we apply them to a social networking system, and a personalized advertising platform through the lens of the developed properties.

3.1 Social Networking

Online Social Network Sites (SNS) have experienced incredible growth over the past few years. It has become an increasingly popular medium for individuals to connect with each other and share a high degree of personal information. From our point of view, an SNS is nothing more than a huge and detailed Identity Management System. This makes such sites a prime candidate by which we can apply the codes that the research has uncovered. Specifically, we will be looking at the Facebook platform.

With over 200 million subjects, Facebook is arguably the most popular social platform today. It has also been the centre of some controversies. Just recently Facebook has been accused of breach Canada's Privacy Laws [22]. More relevant to our considerations, Facebook has recently made changes to the design and flow of the website and has caused backlash among its subjects.

In 2005, Facebook introduced new features that affected the way in which information was distributed to a subject's network on the site. Prior to these changes, information that was inserted or updated on a subjects profile was only visible when the subject's profile page. Facebook then added the *Newsfeed* feature, which essentially aggregated all these information changes and broadcast them to a user's friends. This turned a process from a 'pull' operation to a 'push'. Users reacted against this: Resistance groups were established. The Facebook CEO eventually responded, stating that no privacy options were taken away, and that the information was visible only to the same people who has access as before. *"Nothing you do is being broadcast; rather it is being shared with people who care about what you do"* [4]. Nevertheless, Facebook took down the Newsfeed, and re-released it with various privacy controls.

In their study of the situation, [4] attributed the resistance to individuals' perception of "information access" and "illusory control". Individuals viewed the Newsfeed as increasing the ease with which their information can be accessed by others, and the absence of controls reduced the perceived level of control that subjects had. While this point of view is certainly justified, the properties that have been uncovered here might be able to shed more light on the situation and better relate the changes in the system to the reactions.

The most relevant properties for these scenarios are *control points* and *subject involvement.* Pre-Newsfeed, information was only accessible when the individual's page was visited by another individual. One can technically view this as a single control point. Post-Newsfeed, the number of control points increased dramatically: every person that the information was pushed to represents a control point, where the individual's information is consumed.

In addition, the Newsfeed can be interpreted as a reduction in the level of *subject involvemen*t: In the 'pull' model, visiting an individual's page was a requirement, the page is a representation of the individual on the platform. The individual has taken time to create a profile that represents him/her to others. Therefore, accessing the page can be seen as a control point that has a high level of subject involvement. The Newsfeed represents a loss of involvement, as the information is taken from the individual-controlled profile and to the user at control points that subjects are not aware of or have no control over.

3.2 Targeted Advertising

Targeted advertising has proved to be an extremely lucrative way to increase revenues. This form of advertising involves the tracking of an individual's identity across various services. It could be something as simple as contextual targeting (using keywords based on the content of the current page), or based on individuals' browsing history across one or more sites. These browsing histories and identification details are typically handled in a decentralized manner, making use of cookies stored on the user's computer. These tracking methods have raised issues among privacy advocates.

A recent study found that a significant number of the US population object to the tracking of behavior. Turow et al. [23] found that 86% of young adults reject targeted advertising that tracks behavior across different websites. Advertisers, however, say that individuals - especially the younger generation - do not mind having their habits tracked. Recent developments in targeted advertising have taken the tracking to new levels.

In the UK, Phorm is a company that has developed a targeted advertising platform that is tied directly to a subject's Internet Service Provider (ISP). Every subscriber to the ISP's network is turned into a subject of the system. Every website that a subject visits is passed through the system. It is checked against a list of advertising categories. If a match is found, the category is marked in a cookie and stored on the user's computer. This cookie is then used to provide targeted advertisement on any websites that through the use of a widget. The European Union has recently proceeded with legal proceedings in light of the controversial use of Phorm [24]. The arguments are usually tackled from a high level law based view of privacy rights. Phorm's arguments claim that subjects don't understand the technology and how it works, and that it actually provides anonymity.

Applying structural properties, the items of interest are *subject involvement, discreetness, and the level of control points*. With every website passing through the system, Phorm presents user with a *high number of control points*, resulting in a very restrictive environment for the individual. This situation is exacerbated by *low subject involvement* at the control points: The user's information is taken in a covert manner, without the indiviual being involved in the process. Phorm also provides subjects with a *low level of Discreetness:*. the tracked information is stored on a cookie on the user's computer. In a multi-user environment, the same computer will be used by various individuals that Phorm will not be able to differentiate amongst. When serving customized ads, the system is constantly at risk of revealing a subjects preference by presenting customized content to the "wrong" individuals.

From a metrical standpoint, the properties of interest are *subject coupling, data stability* and *ease of use*. Phorm is a platform used by a user's ISP to deliver targeted advertisements. The relationship between the user and the ISP is that of a consumer paying fees to gain access to the network. This relationship calls for the sharing of certain general and financial information. This is the relevant partial identity of the individual in the subscriber role. By making use of Phorm, ISP's expand beyond this boundary by tracking an individual's habits in depth. This therefore results in low *subject coupling* in the ISP-subscriber relationship. Additionally, an individual's browsing habits are constantly growing and producing a very dynamic data set that results in low *data stability*. In terms of *ease of use*, the system was opt-out, meaning individuals would have to make the effort to request removal from the system.

4 Conclusions

Whilst the use of identity management systems in modern technologies has increased rapidly, the understanding of what constitutes appropriate use of identity lags behind. The disembodiment of modern man from transactions has increased the perceived need to capture the identity of individuals, and developments of systems have largely been driven

by what is technically feasible, and the administrative convenience of the organizations that commission the systems. Whilst the rhetoric of human-centred identity has been plentiful, little research has been carried out to understand the human experience of identity in technology-mediated interactions. This paper presents a first proposal for a set of properties to understand the need of individuals when it comes to identity systems, and what constitutes acceptable use.

The solution proposed here does not aim to replace traditional usability and user acceptance methods: the system properties presented here are complementary. They are specific to identity management systems, and hence help to further explain the potential reactions and strategies that subjects may adopt when confronted with a system. This should help organizations to assess the possible impacts of identity management, and choose properties that meet individuals' needs as well as their own.

References:

[1] A. Giddens, *The Consequences of Modernity*, Stanford University Press, 1991.

[2] D. Lyon, *Surveillance society : monitoring everyday life*, Buckingham [England]; Philadelphia: Open University Press, 2002.

[3] J. Taylor, M. Lips, and J. Organ, *Citizen Identification, Surveillance and the Quest for Public Service Improvement: Themes and Issues*, Joint Session Helsinki, 2007.

[4] C.M. Hoadley, H. Xu, J.J. Lee, and M.B. Rosson, "Privacy as information access and illusory control: The case of the facebook news feed privacy outcry," *Electronic Commerce Research and Applications*, vol. In Press, Accepted Manuscript, 2009.

[5] G. Greenleaf and J. Nolan, "The deceptive history of the 'Australia Card'," *Australia Quarterly*, vol. 58, 1986.

[6] The Register, "Japan rolls out national ID registry," *The Register*, Aug. 2002.

[7] S. Davies, "The complete ID primer," *Index on Censorship*, vol. 34, 2005, p. 38.

[8] WP3, *Study on ID Documents*, Future of Identity in the Information Society, 2006.

[9] M. Lips, J. Taylor, and J. Organ, *Personal Identification and Identity Management in New Modes of E-Government*, Oxford Internet Institute, 2005.

[10] J. Burgoon, "Privacy and communication," *Communication yearbook*, M. Burgoon, ed., Beverly Hills: Sage, 1982.

[11] S.G. Davies, "Re-engineering the right to privacy: how privacy has been transformed from a right to a commodity," *Technology and privacy: the new landscape*, MIT Press, 1997, pp. 143-165.

[12] J.W. Decew, *In pursuit of Privacy: Law, Ethics and the rise of Technology*, Cornell University Press, 1997.

[13] H.J. Smith, S.J. Milberg, and S.J. Burke, "Information Privacy: Measuring Individuals' Concerns about Organizational Practices," *MIS Quarterly*, vol. 20, Jun. 1996, pp. 167-196.

[14] K. Anderson and P. Dourish, "Situated Privaces: Do you know where you mother [trucker] is? ," Las Vegas, NV: 2005.

[15] V. Bellotti and A. Sellen, "Design for Privacy in Ubiquitous Computing Environments," *Proceedings of the Third European Conference on Computer Supported Cooperative Work (ECSCW'93)*, Kluwer, 1993, p. 77-92.

[16] L. Xin, “Trust in National Identification Systems: A Trust model based on TRA / TPB,” Washington State University, 2004.

[17] J. Backhouse and R. Halperin, “Security and privacy perceptions of e-ID: a grounded research,” Galway, Ireland: 2008.

[18] J. Torpey, *The invention of the passport: Surveliiance, citizenship and the state*, Cambridge: Cambridge University Press, 2000.

[19] D. Marks and L. Yardley, *Research methods for clinical and health psychology*, London; Thousand Oaks, Calif.: SAGE, 2004.

[20] U. Flick, *An introduction to qualitative research*, SAGE, 2002.

[21] A. Pfitzmann and M. Hansen, “Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management – A consolidated proposal for terminology,” Feb. 2008.

[22] BBC News, “Facebook 'breaches Canadian law',” *BBC*, Jul. 2009.

[23] J. Turow, J. King, C.J. Hoofnagle, A. Bleakley, and M. Hennessy, “Americans Reject Tailored Advertising and Three Activities That Enable It,” *SSRN eLibrary*, Sep. 2009.

[24] Guardian, “Phorm: UK faces court for failing to enforce EU privacy laws,” Apr. 2009.