**W3C Workshop on Privacy and Data Usage Control**
**Using Network Science To Understand and Apply Privacy Usage Controls?**[1]
**Erin Kenneally**[i]

## Problem and Solution Overview

Information privacy (IP) is an evolution of control tug-o-war between individual rights and interests, social responsibility, and innovation. When users overtly interact with or passively engage a website, social networking or otherwise, they enter a privacy risk zone. Most often, they cannot be sure whether data is surreptitiously being collected, will be used, or further disclosed in ways that contravene their privacy preferences. While we have evolved the 'state of the art' to obligate sites to disclose their privacy collection, use and disclosure policies, protections nevertheless hinge on ex ante user trust that websites will walk the talk.

A privacy solution steeped in information use restrictions and obligations can be approached by better engineering the identification and application of the underlying *reasonable expectations* upon which our privacy controls (laws, policies, standards) operate. I propose that the solution must be derived from the nature of the space creating the problem to begin with, specifically, that a scale-free problem demands a solution derived from scale-free network science.

It seems intuitive that privacy harms are not tied to whether you're at home, or in a park, with a crowd, or in a public phone booth -- yet the trigger for whether we have a reasonable expectation of privacy (REP) is still largely tethered to the contours demarcating public-private spaces. In general, if one's behavior is conducted in 'private' then REP attaches, but if it is exposed to the 'public' then surveillance, tracking, collection, and use of that information is fair game. How do we play the game, however, when our privacy is tethered to information that is decoupled from our persons across the Web ecosystem that does not intuitively fall along public-private boundaries?

We lack a consistent, objective measurement for assessing reasonable expectation of privacy and need to realign standards and their applications to more empirically reflect the contours of the Internetwork environment in which the privacy risks occur and privacy interests need protection. This paper proposes a new way to domesticate REP by using models from network science.

This paper advocates approaching this problem from the conceptual strategy that information privacy is a complex adaptive system.[2] Other legal scholarship has applied this strategy to legal

[1] This paper overviews a recently commenced, applied technology law research project that will result in a more detailed description and evaluation of the utility of scale-free network models to information privacy risk. If the theory is validated, the resulting legal model will offer practical advisement to officers of court, policymakers and privacy risk stakeholders. A version of this paper appears in the American Bar Association Information Security and Privacy Newsletter, Fall 2010.

[2] Complexity, in general, is the science examining the interrelationship, interaction and interconnectivity of various elements within a system and between a system and the environment in which it exists. The hallmarks of complex adaptive systems are distributed

contexts such as environmental policy,[3] intellectual property law,[4] common law jurisprudence,[5] Internet jurisdiction,[6] and information privacy torts.[7] This approach advocates the novel application of network science to a broader value that underpins many of our privacy controls-reasonable expectation of privacy. It suggests that there is a co-evolution between privacy controls (laws, regulations, standards) and informal norms. From this position, REP should be both a top-down and bottom-up tenet, where social norms of what citizens should reasonably expect to be afforded privacy protection should influence our controls, and our controls should shape our REP. As such, an information privacy regime that is predominated by the latter, governance-imposed notion of REP does not objectively reflect the reality of REP 'in the trenches' and threatens to institutionalize a fiction that results in inefficiencies and disrespect for ordering forces that protect individual rights and fosters social good and innovation.

We can infer the incongruity between legacy-driven measures of REP and the changed normative expectations wrought by the Internetwork environment from contemporary controversies surrounding online social networking, geo-location based services, targeted behavioral advertising, and data anonymization. Using a network science model, we may be able to harmonize the understandings and applications of REP across associated privacy controls such as the 4th Amendment, ECPA, FOIA, self-regulatory standards, consumer protection regulations, privacy torts, civil discovery rules, and private contracts and policies. Finally, network science techniques may enable us to operationalize REP into a more predictable, coherent, empirical framework for descriptive evidentiary proof and prescriptive risk management.

---

control, connectivity, co-evolution, sensitive dependence on initial conditions, emergent order, a state not in equilibrium, and a paradoxical condition of both order and chaos. Complex systems are characterized by a large number of similar but independent actors who persistently move, respond and evolve in relation to each other in an increasingly sophisticated manner to generate emergent order. See, e.g., JOHN H. HOLLAND, HIDDEN ORDER: HOW ADAPTATION BUILDS COMPLEXITY (1995); MITCHELL RESNICK, TURTLES, TERMITES AND TRAFFIC JAMS (1997). Cited from Matwyshyn, Andrea M., Organizational Code: A Complexity Theory Perspective on Technology and Intellectual Property Regulation. Journal of Technology Law & Policy, Vol. 11, No. 1, 2006. Available at SSRN: http://ssrn.com/abstract=914783.

[3] See, e.g., Daniel A. Farber, Probabilities Behaving Badly: Complexity Theory and Environmental Uncertainty, 37 U.C. Davis L. Rev. 145 (2003) (using power law and other complex systems to evaluate different environmental law models for dealing with uncertainty).

[4] See, e.g., Andres Gonzalez, Change On The Creation, Dissemination, and Protection of Intellectual Property: Scale-Free Law: Network Science and Copyright, 70 Alb. L. Rev. 129.

[5] Smith, Thomas A., The Web of Law (Spring 2005). San Diego Legal Studies Research Paper No. 06-11. Available at SSRN: http://ssrn.com/abstract=642863 or doi:10.2139/ssrn.642863.

[6] Matwyshyn, Andrea M., Of Nodes and Power Laws: A Network Theory Approach to Internet Jurisdiction Through Data Privacy. Northwestern University Law Review, Vol. 98, 2004. Available at SSRN: http://ssrn.com/abstract=904074.

[7] Strahilevitz, Lior, A Social Networks Theory of Privacy (December 2004). U Chicago Law & Economics, Olin Working Paper No. 230; U of Chicago, Public Law Working Paper No. 79. Available at SSRN: http://ssrn.com/abstract=629283 or doi:10.2139/ssrn.629283.

- **Problems With The Old Playbook**

Thresholds for privacy protections are anchored around the principle of `reasonable expectations,' which explicitly or implicitly underpins institutional privacy controls- laws, regulations, industry standards, private agreements. The incumbent standard for measuring and applying REP is obsolete and contemporary cases and controversies shine light on our need for a new model. This legacy is largely anchored in the delineation between public and private spaces (closely related is the third-party doctrine[8]), which broadly holds that what one knowingly exposes to the public loses any expectation that it is deserving of protection under the law.

We have defined what public-private means with respect to a person's location, property and behavior, but with regard to his information in an online social network space, the legacy metrics are exposed by unresolved cases and controversies as too coarse and unreflective of the nature of the threat to the underlying privacy interest. Our privacy expectations --the interests and rights associated with relationships between persons with respect to the collection, use and disclosure of data -- are informed by controls (laws, policies, standards) whose measurements and applications of privacy are ill-fitting given the information privacy threat model.

- **Heeding Fractures in the Playbook Design**

The incumbent REP approach may be faulty because it lacks acumen of the conditions of information collection, use, and disclosure that reflect the Internetwork structure comprising the playing field for information privacy. The current REP paradigm is fundamentally contoured around public versus private measurements that presume a *scaled* network of information flows where every PIA controller (PC)[9] is equivalent. It predominantly treats all disclosures to third parties identically rather than framing privacy risks empirically according to the *fitness* of and *scale-free* relationships between PC, and thereby sets the stage for incongruous protection and enforcement of rights.

- **A New Playbook Using A Scale-Free Network Model**

The social network environment demands an evolved privacy risk management model that correlates to the contours of this new context. Relative to the legacy, offline context of privacy there are certain features of the online setting that render legacy metrics for determining REP unsuitable: there is much less awareness and understanding of the technology underpinning PIA location and movement; the data relevant to privacy interests is continuous and as such privacy choices are not discrete and linear; and, the boundaries that inherently define privacy

---

[8] For an overview of the application of the third party doctrine to 4th Amendment Law and information see, e.g., Henderson, Stephen E., Beyond the (Current) Fourth Amendment: Protecting Third-Party Information, Third Parties, and the Rest of Us Too. Pepperdine Law Review, Vol. 34, 2007; Widener Law School Legal Studies Research Paper No., 08-11. Available at SSRN: http://ssrn.com/abstract=922343, Lawless, Matthew D., Third Party Doctrine Redux: Internet Search Records and the Case for a 'Crazy Quilt' of Fourth Amendment Protection. UCLA Journal of Law and Technology, Vol. 11, p. 1, 2007. Available at SSRN: http://ssrn.com/abstract=984314.

[9] PIA Controllers refer to the persons or entities – third parties or data subjects themselves – who possess and have the capability to disclose PIA. This label intentionally avoids any reference to ownership interest in PIA.

are now virtual and not sentient, thereby rendering privacy risk more opaque.[10]

In order to move beyond the circular paradigm where our privacy controls apply REP by what is deemed 'private' and vice versa, we need a model that organically measures REP based on the capabilities and risks in the cyber environment, and one that allows those new metrics to inform a reshaping of our privacy controls. Modeling information privacy as a scale-free network may enable a more tailored understanding and prescribing of the conditions necessary for a privacy control regime to succeed. Stated differently, if privacy risks and interests flow with the PIA vis-à-vis a scale-free network, this model promises to inform a better application of privacy controls.

Scale-free network playbook basics

Networks are comprised of nodes connected by links. Lay examples include: scientific research paper citations, the sales of books and branded commodities, the World Wide Web, webpage hits, the number of citations on scientific research papers, and certain criminal activity. If information privacy is a scale-free network,[11] it exhibits some fundamental characteristics:

1. The distribution of nodes approximates a power law distribution, where few nodes have many links (aka, hubs) and the majority of nodes have few links.[12]

2. The network evolves and is dynamic, meaning that nodes are added and removed throughout time.

3. Links exhibit preferential attachment, commonly coined as 'the rich get richer,' whereby new links are added to nodes based on either the number of existing links or some measure of fitness of the node.

Scale-free network playbook for information privacy

Hypothesizing that we can model information privacy according to the properties of a scale-free network in order to better describe and protect REP means that the nodes are the PC (PIA Controllers)[13] and links are the collection/disclosure of PIA (the 'citations' of PIA). PIA is connected by links between PC.

- **Operationalizing the Scale-Free Privacy Playbook**

Before we can operationalize an evolved assessment of REP to inform privacy controls using network science techniques, ensuing research needs to validate this approach by exploring the following types of questions:[14]

* Is information privacy a scale-free network? If so, what does it mean for describing and prescribing REP? For example, what are the possible normative implications for information privacy law, such as whether PIA exposure to 3rd parties is a de facto poor indicator of greater threat to privacy? How might knowledge of PIA flows either eliminate the use of public-private standard for measuring REP; or, can it be used to re-define what we mean by public-private

[10] Camp, Jean. "beyond consent: implications of ubicomp for privacy"

[11] This is juxtaposed to a scaled, random network that follows a bell curve where nodes have a typical number of links. For an authoritative overview of the theory, see ALBERTO LASZLO BARABASI, LINKED (2002).

[12] This concentration of nodes is highly skewed.

[13] Insofar as the PIA is expressed and exists via its PC, the nodes also represent PIA.

[14] See, Matwyshyn and Smith, supra notes 7 and 8.

space with a fidelity that is more aligned with the reality of information flows? How well are certain PC integrated with the whole system, such as data aggregators or online advertising networks? How closely does the geo-location of PC hubs correspond to traditional public-private and 3rd party doctrines?

* How should we apply a scale-free model to privacy controls? For example, does knowing how PC ages enhance our understanding of how privacy evolves with time? Can the PC churn rate help us understand how quickly PC accumulate links and determine the rate of collection/disclosure of PIA? Should the size of PC clusters and their proliferation establish living REP or indicate failure of privacy controls?

* Is there congruence between collection/disclosure topology and the semantic topology of PIA? For example, do the clusters of PC link based on shared meaning of the value of a particular PIA for price discrimination or some other economic use?

Subsequent to validating the aforementioned foundational assumptions, we can operationalize REP to achieve the following intended benefits:

* Inform evidence-based policymaking -- ensure that choice and control of the collection, use and disclosure of PIA is based on empirical reality of how it flows throughout networks; inform default privacy presumptions in an effort to devise more efficient contractual rules, e.g., should we impose implied nondisclosure obligations on certain PC for certain categories PIA? Or, should privacy settings or terms of service establish default REP in web communications?

* Enable better privacy risk management for both individuals asserting privacy rights and entities handling PIA – the entities with countervailing interests—through more predictable outcomes, more certainty about REP determinations, and lower liability risk.

* Advocate common definitional semantics to harmonize reasonable expectations across privacy controls- industry-specific and data-specific laws, geopolitical authorities responsible for enforcing privacy controls, and between and among industries that are largely privacy self-regulated.

* Refute or validate non-institutionalized intuitions about REP norms.

* Devise more sophisticated justifications for our intuitions about privacy (e.g., autonomy, seclusion, property).

- **Conclusion**

Information privacy that follows a power law distribution means that few PC have many collection/disclosure links and most PC have few collection/disclosure links (imagine a long tail distribution diagram, a right-facing hockey stick, or a left-to-right down-sloping line on a graph). Network growth means that the PC nodes are dynamic and unstable, and the PC and PIA is altered, removed, and added. To say that the links are not invariant means that PC with many PIA links will generally increase collection/disclosure links more and/or they will preferentially attach because they are more fit than other PC (e.g., they have more valuable PIA, their links are more authoritative, etc.).[15]

[15] From this perspective, the old privacy context which followed traditional bell curve averages, most PC had about the same amount of PIA, PC were relatively known, stable, and static; and, links were relatively clear, predictable, and uniformly distributed. The major implication here is that it was a context that was more conducive to binary REP assessment and privacy controls applications.

[i] **Author Bio**

Erin Kenneally is a licensed Attorney and Forensic Scientist who advises, researches, publishes, and speaks on prevailing and forthcoming issues at the crossroads of information technology and the law. These include legal, technology, and policy implications related to information forensics, information security, digital evidence, privacy technology and information risk. Ms. Kenneally is founder and CEO of Elchemy, Inc.; holds a Research Specialist position at the Cooperative Association for Internet Data Analysis (CAIDA) and the University of California San Diego; and, serves on private and public sector advisory bodies. Ms. Kenneally holds Juris Doctorate and Master of Forensic Sciences degrees.