Revisiting Signals and Noise for Ethical and Legal Research Using Online Data

Erin Kenneally
University of California San Diego
9500 Gilman Dr.
San Diego, CA 92093 USA
erin@caida.org

Abstract
The stability of trust on the Internet has implications for political diplomacy, innovation, economic stability, social and civil relations, and individual self-determinism. The degree of online trust is a reflection of the gap between individual and collective Netizens’ expectations formed by laws and ethics, and their capabilities enabled by technology. Law and ethics, just as with familiar offline society, act as ordering forces that inform the acceptability of our behaviors and relationships with other person and organizations. The migration of these analog activities online has exposed a rather sweeping gap between expectations and capabilities, where legal and ethical ordering forces are challenged to re-examine, interpret, and apply the tenets and principles upon which they moor. As this gap widens, so too does ambiguity between asserted rights, interests, and threats to same.

This gap is manifest most prominently in the current industrial and geo-political struggle to define rules of engagement for cyber conflict and national security, as well as with online advertising and data brokering. A related context where ordering forces are challenged, lower on the public notoriety index but no less considerable, is information and communication technology (ICT) research. The controversy over the collection, use and disclosure of online data for
research exposes gaps and deficiencies in the legal and ethical structures that directly and indirectly inform and reflect our expectations.

Notably, “consent” has been a fundamental mechanism for protecting rights and interests in both law and ethics. As such, it serves as an institutionalized signal for persons’ reasonable expectations. Yet, the ability to easily collect and combine massive amounts of existing, “publicly-available” information of a sensitive nature (personal or confidential) online exposes deficiencies in consent as an effective signal for expectations.

More specifically, researchers increasingly encounter data online such as personal health, financial or behavioral records; usernames and passwords lists; corporate manuals and technical documents; email and voice communications databases; and, device vulnerabilities and machine-to-machine communications. It is located in various online locations ranging from normal websites and social networks to underground criminal forums, Internet relay chat rooms, and publicly-obscured/hidden sites. And, its availability is often a product of malicious, negligent, or ignorant collection or disclosure by a third party.

In this context, consent as an expectation signal is strained along substantive and procedural dimensions. For example, some argue that the existence of other signals (i.e., the data was public and/or non-identifiable, the purpose of the research is to study a system or threat and not individual persons) pre-empts the need for consent. In addition, obtaining consent for what amounts to secondary use of online data may be impracticable in light of the distance between researchers and data subjects, or outweighed by countervailing intended benefits or academic freedom interests.

With research using data available online, researcher conduct is not fully prescribed or proscribed by formal ethical codes of conduct or law because current expectations signals are ill-fitting. This presentation is intended to advance the collective dialogue toward a path that revisits and harmonizes ethical and legal signals for research using online data among researchers, oversight entities, policymakers and society. It does not dictate answers but aims to point out where current ordering forces breakdown in the context of online research and to suggest how to identify and respond to these grey areas by applying common legal and ethical tenets.

**Author Keywords**
Ethics; Law; Technology Policy; Research Ethics; Risk

**ACM Classification Keywords**
K.5.0. LEGAL ASPECTS OF COMPUTING: General.
K.4.1. Public Policy Issues: Ethics; Privacy; Use/abuse of power.