W3C Workshop on Privacy and Data Usage Control

UCDC: A User-Centric Disclosure Control Solution To Privacy Interoperability

Jim Lookabaugh jim@intergreen.com

Erin Kenneally erin@elchemy.org

Problem Statement: *Stakeholder Tensions*

Whether the commercial value of personal information is a result of discovery or design, it has unmistakably catalyzed a nontrivial tension between service providers and citizen-consumers. As such, personal information resides at the crosshairs of market capitalism, innovation, individual privacy, and social good. In this space, the value of personal information as a commodity has transformed privacy beyond a zero-sum game and into a risk-benefit balancing act between stakeholders.

A privacy solution around data use and handling must account for the needs and risks of both the users and the service providers. Often, prevailing solutions are dependent on ex post facto enforcement of data use policies. These solutions have data as typology-centric rather than data as flow-centric. And, these solutions do not capture and/or carry forward the relationship between privacy elements: use purpose, user, data type, privacy-relevant action, obligations and conditions. Therefore, invariably the enforcement of privacy preferences and mitigation of risks is predicated on being able to granularly express them in ex ante policy that can account for possible data flows.

Consumer-users of social and other Internet services want informed choice and control while still availing themselves of the ever-growing service offerings, and demanding a low barrier to achieving both desires. Service providers are incented by revenue streams to commoditize personal information in ways that may contravene user choice and control, including externalizing personal information across the heterogeneous service provider environment. Yet, they are obligated to respect those desires both by formal law and regulation as well as by realizing the value of privacy in attracting and retaining customers.

Gap in Solution Space

Amidst the confluence of multifaceted demands to balance tensions, current privacy-related solutions in the marketplace have shifted to address the problem of interoperability. Much effort has been invested in approaches where policy rules may be persisted and managed by one party, transported to another party for automated evaluation and decision-making, and often further transported to yet another entity for automated enforcement against the pertinent information at hand. While the underlying machinery responsible for automating the policy interpretation and execution certainly rely on these approaches, they have yet to gain widespread adoption across social networking sites both from a provider and user perspective. Where interoperability has gained traction, often it has taken

shape as a partial privacy solution that addresses the conceptual and operational space of access control, thus failing to address the larger risks from use and disclosure.

Regardless of the architecture chosen for interoperating the privacy policy management and enforcement across providers and users, integration of privacy enhancements into existing data handling tools is further challenged. First, inbound data must be prepared for any permutation of a set of data elements, because variances in policies across the stakeholder population may dictate the presence and absence of elements in a myriad of ways under an enormous set of circumstances. For many, this may mean becoming more flexible on schema specification. Second, domain-level semantic savvy is required in order to implement machine understanding of the inbound data, since its format and structure may be difficult to anticipate as described above. These challenges are rarely addressed in today's panoply of current data-handling tools and solutions marketplace.

We suggest that a necessary prerequisite to propagate machine-level interoperability between the transport of privacy restrictions and obligations among providers is a pragmatic, comprehensive and user-friendly way to express and manage privacy and disclosure control policy by stakeholders responsible for drafting policy. Policy begins with people's domain-knowledge and informed expectations and ends in the technical implementation of policy enforcement, yet the gap between conception and implementation has impeded cross-provider flow of preferences with accompanying data.

A Disclosure Control Solution (UCDC) – Under the Covers

Information privacy solutions involve the control of data usage and the control of data handling, but these are predicated on the control of data disclosure. Data disclosure is the concern of specific data in a specific context (i.e. in association with some other data) that may be possessed, no matter how briefly, for a given circumstance. It is a space conceptually different from that for access control.

• Recognizing the Relationship Between Access Control and Privacy

Access control outcomes are often defined as the set of Permit (Grant), Deny, Indeterminate, and Not Applicable, and this set may benefit from extensions specific to the privacy and information disclosure control landscape. Disclosure control outcomes may be characterized as discrete but more subtle than access control outcomes. This outcome set may include Not Applicable, Disclose (Permit/Grant), Redact (Mask), Withhold (Deny), Hold For Review (Indeterminate), and Don't Disclose Out Of Context. This final member of the outcome set regards the declared prohibition on piecemeal redaction/withhold of hierarchical data yet allows wholesale redaction/withhold. For social networking privacy, the authors' solution is targeted at disclosure control and is called UCDC, which supports such an outcome set.

Upon granting access or disclosing information, the resources now in command of the user may be subject to certain covenants or obligations to which

the user is held by the granting/disclosing party. Examples include exploiting a resource only for official, justifiable business purposes, retaining disclosed information for a maximum period of time, and prohibiting dissemination to other parties. UCDC and many access control solutions are designed to support policy-driven obligations.

• Allowing Both Service Providers and Users To Set Policy

Anyone with a legitimate stake in the appropriate disclosure of information ought to be able to define policy rules. Therefore, the solution must be easy to learn and use by each stakeholder. The solution also must support the concept of managing the legitimacy of members in the policy-making stakeholder communities.

The user should be able to define and reuse one's own policy, because the user is the sole member of one legitimate stakeholder community. There are other such communities, such as the community of service providers. Negotiation can be accomplished through publishing one's policy for all to read, but this is not an approach to automated negotiation. By knowing what service offerings are dependent on the presence of certain disclosed data items, a user reading the service provider's policy will be empowered to make decisions about what data to keep private at the expense of sacrificed service offerings. If the service provider wished to read the user's policy (or better yet, to survey a representative sample of all its users' policies), the provider can better understand each user's concerns and better understand the perceived value propositions of its services by the user community.

• Encouraging Users to Define Privacy Policy Rules

The perspective of service provider as a good citizen in actively encouraging users to define privacy policy is based on some assumptions. First, users need to be educated concerning the need to define policy and concerning the exposure to various risks in the absence of defined and enforced policy. Second, user-adoption of a solution to define policy necessitates that the solution be not difficult to learn or to use, which dictates that the policy language must be easy to read and to write, easy to accurately comprehend, and that the policies must be easy to manage. Policy languages targeted for machine interpretation are typically inappropriate for human use. So, an additional language to a machine-centric solution or a different solution altogether is required. UCDC offers a language specifically designed for humans to express policy.

Third, users will appreciate a solution that can be leveraged at the user's discretion across services, across systems, and across domains (e.g. across social networking, health care, and banking). Users will gravitate toward a solution that allows policy to be written once and enforced under many scenarios. However, the solution must be attractive enough to users in order to convince corporate decision makers that the education costs are justifiable and, a more user-friendly solution will be easier to sell.

• Guaranteeing that Data Use Matches Its Policy-Driven Purpose

Data handling guarantees are impractical in a distributed storage/processing environment, where non-authoritative systems have historically been permitted to persist copies of data retrieved from authoritative sources. These guarantees are also impractical in a non-uniform political environment, where a universal outcome, such as honoring a data retention horizon, cannot be expected among multiple parties/jurisdictions. Some parties can be expected to exercise their sovereignty and to respect their own critical requirements at the expense of others' requirements. The solution can guarantee two things. The first is that the service providers and the users can be enabled with the means of expression of policy (at least to a certain range of levels of complexity). The second is that policy will be available to be retrieved by the resident process at the point of data disclosure/usage (at least from a software system standpoint). Neither ignorance of policy nor inability to express policy rules can be an excuse then for inappropriate data disclosure. Both of these guarantees are part of the design approach of UCDC.

• Recognizing the Role of Semantics

The intersection between policy dependencies on the data at hand and the dependencies on how that data is characterized (its metadata) relies on resolving common values. Successful resolutions of commonality in this intersecting space are critical in a privacy solution or any information disclosure control solution.

Data values, data types, and data labels (e.g. XPath expressions) have historically been used as the basis for resolving commonality. However, policy authors may not have familiarity with data types or labels, and those types and labels may change as schemas are improved over time. This casts doubt over the efficacy of user reliance on citing dependencies of policy on data types and labels. Furthermore, policy will frequently be independent of data values, considering the variability of values for a given datum and the burden it would be to draft policy for every possible value in a datum's range.

Therefore, this intersecting space will benefit both policy authors and data processing engineers by an additional reliance on semantics. As a semantic glossary expands, the support grows for describing additional rights, obligations, data, conditions, and even the decision outcomes themselves. A semantic ontology – or more likely several ontologies where each is managed by a discrete business domain or jurisdiction – will be helpful.

The use of references, such as web-based URLs, to link together semantics within and across ontological domains may minimize needs to establish standard semantics for concepts, because multiple semantics can be easily linked to each other as being synonymous. For example, a semantic authority may decide at some point to link a semantic reference of "earlobe pulse oximeter" to another of "obsolete medical equipment", rendering from that point forward all data semantically known as the former to being known also as obsolete equipment.

• *Semantically Interoperability Across Distributed Environments*

Certainly, many of those who write policies, particularly the users of social networking services, will have no familiarity with technical specifications of data

exchange schemas or data formatting requirements. Therefore, such policies will lack dependencies on those technical aspects of the data. It can be argued that those who engineer the software systems and the data exchange schemas ought to have knowledge of the business domain jargon used by policy authors, but in an ever increasingly integrated landscape of disparate business domains, it may be increasingly difficult to statically imbed such business domain knowledge within these system implementations, let alone the difficulty to acquire all that knowledge.

It is imperative for a successful privacy solution to bridge these two worlds: the world of policy and the world of data processing. To do so, it may be beneficial to engage a third world where semantics are managed for shared use by policy authors and by data processing engineers. As such, tooling used by occupants of the policy world and of the data processing world will need to facilitate the exploration and discovery of relevant semantics that are managed in this separate space.

There may be no means by which to guarantee the same or equivalent semantics are used by policy authors and by engineers. Anticipating that discrepancies may exist, it would be prudent for the policies to declare conservative behavior by default and to more strongly preserve privacy / to more strongly constrain disclosure. Concerned parties who observe that disclosure is too constrained can then engage relevant stakeholders to take remedial action by refining the usage of semantics in each space or extending the semantic ontologies as appropriate.

UCDC and many access control solutions permit policy authors to declare default behavior when one or more rules cannot be positively asserted to apply to data. Yet UCDC goes further in being semantically savvy by design, which supports immediate recognition of remedial refinement in semantic ontologies.

• UCDC and Privacy of Social Networking Users

UCDC is a solution model for the particular concerns of privacy preservation for users of social networking service providers. UCDC fosters adoption by offering a policy expression language that is easy to read and to write for humans. It intentionally minimizes conventions and symbols that could facilitate machine interpretation/processing but that only too often serve to dissuade humans from using it.

It satisfies needs of users and service providers by allowing both of these stakeholder communities to manage themselves, honoring division of responsibility. It promotes confidence by its design that only legitimate stakeholders' policies will be evaluated and enforced with a reliable reconciliation of contentious outcomes. UCDC is semantically savvy and frees users from knowledge of internal details of message schemas. It frees providers' data handling engineers from knowledge of users' domain jargon. UCDC helps complete the puzzle and finally solve privacy concerns in social networking by being focused on the privacy and disclosure control space, supporting a broader set of outcomes than simply grant or deny.