8677 Villa La Jolla Dr., #1133
La Jolla, CA 92037
United States
E-MAIL admin@elchemy.org
CYBER: http://www.elchemy.org

To: ICANN Expert Working Group
From: Erin Kenneally, CEO, Elchemy
Date: 5 September 2013
RE: Input / Questions Regarding the Next Generation gTLD Directory Services Model

These recommendations are best described as addressing the risk analysis process and model that necessarily must coordinate the next generation Registration Directory Service's desired features and design principles with information risk controls, within various collection and disclosure scenarios. As such, the approach suggested is responsive to various aspects of the EWG's questions related to the Purposes and Users (including Use Cases), and Requirements for Data Elements, Privacy, and Access and Accountability components, most especially.

As Co-Principal Investigator for a multi-year effort[1] that has produced a practicable Disclosure Control Framework (DCF)[2] for collecting and sharing computer network and security data, I contend that this model is quite well-suited to address the fundamental utility and privacy challenges that the new WHOIS aims to overcome. Given the competing interests of various individual and organization stakeholders that must be considered for the ARDS, a disclosure model that frames solutions as a positive-sum proposition is prerequisite to achieving extensible and defensible decisions.

My primary critique of the draft ARDS model is that while the EWG has done quite a commendable job articulating impacted Users and related needs (Purposes & Use Cases), it lacks a mechanism for identifying, evaluating and balancing the associated risks that arise when those desired outcomes are viewed from an ARDS systemic perspective. This is manifest, for example, in a host of questions raised by the EWG related to data element risks such as which requestors should have access to what data elements. In other words, in order to pragmatically satisfy accountability, privacy, and operational needs of stakeholders, the work-in-progress ARDS design should be guided by an overarching framework that coordinates the features with risks.

As such, the DCF may be instructive. The DCF outlines a process involving three phases– risk and utility analysis, application of controls and assessment (Figure 1), that maps rather appropriately to the User-Purpose-Use Cases already outlined by the EWG. Each phase in the framework discusses, in necessarily general terms, the primary considerations essential for all ARDS data sharing scenarios: the desired utility objectives, relevant risks, options for disclosure controls, and the impact of those controls on the chosen risk and utility determinations. Notably then, the DCF allows the existing ARDS components to be viewed and operationalized into a workflow that considers data risks inline with functional goals, all within a range of collection and disclosure control options.

---

[1] This applied R&D project has been supported by the U.S. Dept. of Homeland Security, Science & Technology Division to help advance the Protected Repository for the Defense of Infrastructure Against Cyber Threats (PREDICT) program, a centrally-organized but distributed repository of data for government, academic, and industry operators and researchers (www.predict.org).
[2] Coull, Scott E. and Kenneally, Erin E., *A Qualitative Risk Assessment Framework for Sharing Computer Network Data* (March 31, 2012). 2012 TRPC. Available at SSRN: http://ssrn.com/abstract=2032315 or http://dx.doi.org/10.2139/ssrn.2032315.

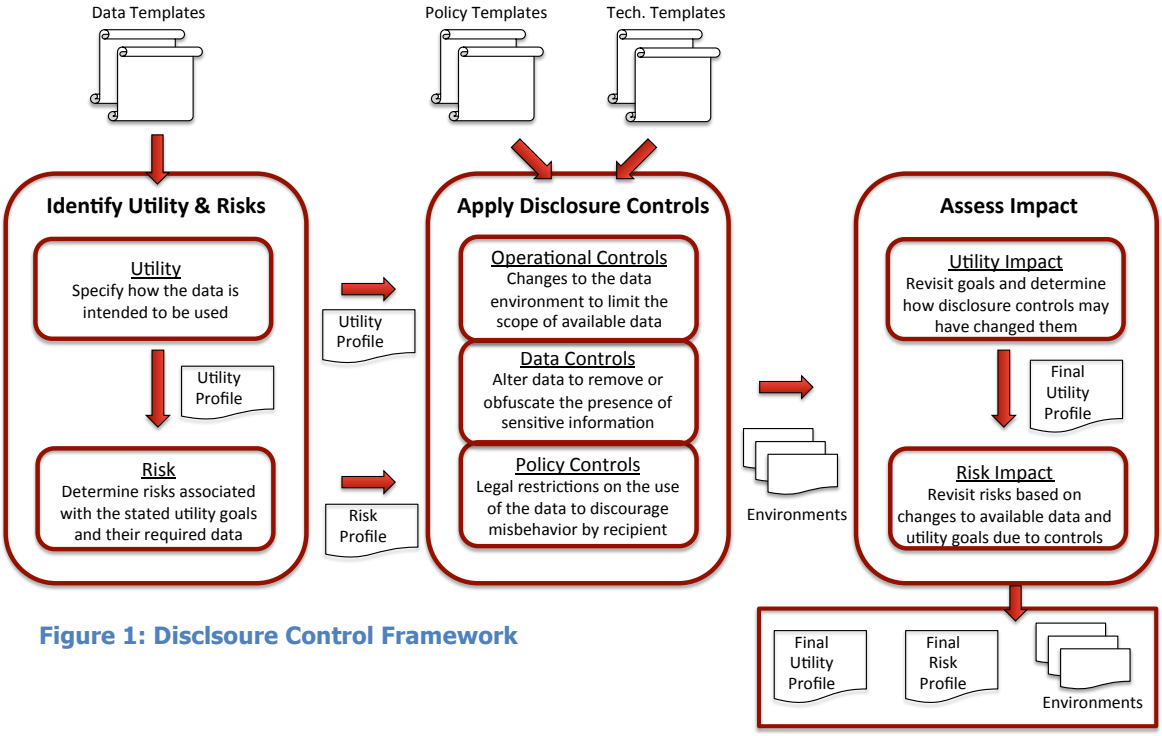

Figure 1: Disclsoure Control Framework

While a detailed mapping between the existing ARDS components and the DCF is beyond the scope of this comment, I highlight below how the ARDS can be conceptualized according to the three DCF phases. In the first phase utility identification, the ARDS as data publisher would plug in the various use cases (users and purposes). The DCF normalizes "utility" along the following dimensions:

- **Audience**: the intended User-recipient of the data. The audience continuum can be roughly divided into individuals at the most restrictive end, consortiums in the middle, and public release at the most relaxed end.
- **Duration**: how long can the User-recipient access that data. Data access durations may span from short-term access for real-time operational uses to indefinite access for general research purposes.
- **Timeliness**: how quickly the data is made available to the User-recipient. The continuum of setting choices ranges from real-time access to longitudinal data collection with long lag times between collection and availability.
- **Detail:** the level of detail required by the User-recipient. Some operational tasks require detailed data about events or records, while general research use is most concerned about overall trends that manifest themselves in the data. Therefore, the ARDS may consider a continuum of data detail from event-specific data to general trend information derived from the data.
- **Functionality**: the level of specificity associated with the use cases. This may span the spectrum from data releases tailored to specific concrete tasks, to the other end involving open-ended tasks.
- **Output**: the intended outcome of using the data, or the output of the User-recipients' interaction with the data. This can run the gamut between private knowledge at the most restrictive end of potential outputs, to publication at the broadest end of dissemination.

As for the first phase risk identification, it is comprises three properties that capture the relevant information necessary for the ARDS to explicitly describe the risks and their sources. These properties include:

- **Type of Data**: The type of data involved in the given use case may impose various

obligations or restrictions related to the standard of care in collecting, using or disclosing it. For instance, Personally Identifiable Information (PII) has specific restrictions on its use imposed by laws and ethical risk sources. Mapping out the specifics types of data, from among all fields available in the relevant data templates, helps to identify the pertinent risk factors.

- **Participants**: The role of the characteristics of the ARDS as publisher and various User-recipients also alter the impact of the risk factors. Regarding the User-recipient, their overall potential for abusing the disclosed data should be considered, including motivating factors (e.g., economic advantage, notoriety) and their technical expertise in bypassing any applied disclosure controls.
- **Risk Factors**: These are the actual identified risks that arise due to the combination of participants and data for the given use case. Each risk factor is derived from risk sources that include laws, private agreements, proprietary rights, ethical obligations, unilateral policies (e.g., Terms of Use), and best practices. The combination of data and participants should be evaluated against each of the categories for the data sharing scenario in question, and relevant risk factor should be listed with their source.

The second phase of the framework considers the Phase 1 risk-assessed data and utility-ascribed data in parallel and offers a menu of disclosure control options for the ARDS to apply to the shared data to achieve appropriate risk and utility since each option has measurable impact on each of the two components. These controls are organized as **operational**– how the User-recipient may interact with the shared data (e.g., filtering, format encoding, access such as query interface or bulk download), **technical**– how the data can be altered to prevent sensitive data leakage (e.g., aggregation, pseudonymization, deletion), and **policy**– how the ARDS can address the identified data risks ex post via contractual and policy-oriented agreements concerning the access, use and secondary disclosure of the data by the User-recipient.

To the extent that the ARDS may be designed as an implementation of the DCF, the above risk assessment framework could be operationalized[3] with the addition of **templates** - data structures that encodes information about disclosure control components (data elements, policy, technology), and **environment**s- a concrete instantiation of the disclosure controls and related data sharing infrastructure and chosen by the ARDS for a particular use case scenario (e.g., access controls, server software, etc.).

Thank you for your work on this pressing issue and effort, and for your earnest consideration of what I assert can help the ARDS become as effective as possible.

Respectfully Submitted,
Erin E. Kenneally, M.F.S., J.D.
CEO, Founder Elchemy, Inc.
Erin@elchemy.org

[3] Coull, Scott E. and Kenneally, Erin E., *Toward a Comprehensive Disclosure Control Framework for Shared Data*, publication forthcoming, 2013 IEEE International Conference on Technologies for Homeland Security, Boston, MA (Nov 2013).