

## Effective Practices and Solutions

Title: *Data Categorization and Minimum Security Standards*

Background: Cornell has existing university policy that assigns the responsibilities of data security and confidentiality to a set of nine data stewards. These data stewards are charged with determining how to manage, access and utilize the data they are responsible for in a manner that is consistent with university policy and does not jeopardize the security and confidentiality of those data. While in principle this policy take a solid stand toward data security and protection and was able to get wide support, in practice this has fallen short in many area. Many data stewards have not taken the necessary steps to define the sets of data they are responsible for, have not outlined requirements and processes for data access and the policy has lead to inconsistencies in implementation. A particular challenge is found across data sets that are shared among several data stewards such as social security numbers.

An additional complicating factor with data protection is the changing or emerging security requirements, expectations and risks. Due to these changes the university must enact wide changes in data management for access and collection of future data but, and maybe more importantly, must also clean up our IT environment to adequately protect older data under new requirements. The bottom line is that university data have not been sufficiently managed. A good example is the situation with social security numbers. Prior to about five years ago, the university used social security numbers for almost everything to include registering students, posting student grades, identifying employees, in all HR records and even to climb our rock wall. With the introduction of new rules and expectations concerning the appropriate use and confidentiality of social security numbers the university did not only need to change our practices but also needed to clean up all social security numbers that remained from previous practices. This has not been an easy process given the wide use and distribution of confidential data such as social security numbers. When we began this exercise it was estimated that approximately 60 percent of Cornell's administrative computers stored confidential data – typically unacknowledged by the primary system user.

Description: Cornell has written a data classification and security policy that is to be complementary to the existing data stewardship policy. This new policy will require the established data stewards to clearly define and map the data they are responsible for into one of three data classifications and system administrators/owners to maintain their computer systems according to a set of established security configurations.

The established data classifications are *Confidential*, *Restricted* and *Public*. While the task of mapping all data is a daunting one, Cornell has established a default category of *Restricted* for all data unless otherwise categorized by the data stewards as *Private* or *Public*. This strategy not only serves to streamline the data categorization process but also allows Cornell to ensure all administrative computers are configured to meet reasonable security standards. The benefit of this is easily demonstrated through an analysis of past computer compromises that shows virtually all of the compromises could have been prevented if the requirements for *Restricted* data had been followed.

As an initial starting point identity theft data (social security numbers, credit card numbers, driver license numbers and banking information) and HIPAA data were

mapped into our *Confidential* category. Data stewards will then map additional data they are responsible for into the *Confidential* or *Public* categories as they feel necessary.

To successfully develop this policy, the IT Security Office worked very closely with the wide campus IT community to establish a set of requirements that could be met and followed. It was extremely important to have IT leadership support. The Security Office must also continue to work with the data stewards to help them understand both the advantages and disadvantages of data classification is accomplished (higher security means additional costs and additional process and/or oversight).

The establishment of the data security policy has served to highlight gaps in our current Data Stewardship policy that need to be addressed. Regulated data such as FERPA, HIPAA or PII that are used across a number of data stewards are not adequately or consistently being addressed.

Finally, data discovery is a very important aspect of this work. Over the years, we have been less than strict with respect to how are data were managed, protected and tracked across the institution. This has resulted in a vast spread of data across the institution and most people not realizing the type of data on their computers – we can't protect unless we know the data are there. Cornell has invested significant energy in developing an application called Spider that can be run on computers to more efficiently search for social security numbers and credit cards. Many Cornell units, state and national governments and worldwide organizations have used this as a mechanism to clean up old, unnecessary data.

#### Benefits:

1. Establishing consistent and meaningful data categories permits clear requirements, cleaner user awareness and more thorough auditing
2. Using a default category means we don't have to work through all of our data to begin implementation
3. Establishing a baseline for all IT means bringing up our overall security posture considerably resulting in reduced computer compromises
4. Wide acceptance and support from the IT community due to requirements development process
5. Consistent procedures and process accepted by the data stewards
6. Established security requirements can change as technologies or threats change

#### Shortcomings:

1. Need to ensure data stewards categorize their data appropriately. We need to resist the urge to move all data into the highest security classification
2. Data that are shared between data stewards will still be challenge. For example, SSNs fall across several data stewards but require consistent procedures
3. The current policy only addresses data that are stored in electronic form – other forms of information such as paper must also be addressed
4. Compliance assessments will be difficult

#### Implementation Challenges:

1. Working through full data categorization with our data stewards

2. Consistent implementation of the requirements
3. Awareness and training needs
4. Finding all of our data

Future Plans: We will continue to update and strengthen the security requirements as technologies are introduced and threats changes, common practices and procedures need to be established and non-IT data must be addressed.

References: *You may list up to five complete public URLs and attach files to provide more information about your practice/solution (e.g., campus reports, a Web site supporting the practice/solution, a paper or presentation describing the practice/solution in more detail). Add any attachments necessary to further illustrate your practice (e.g., PDF, Word document).*

Data Stewardship and Custodianship policy

[http://www.policy.cornell.edu/CM\\_Images/Uploads/POL/vol4\\_12.html](http://www.policy.cornell.edu/CM_Images/Uploads/POL/vol4_12.html)

IT Security Requirements

<http://www.cit.cornell.edu/security/requirements/>

Draft Security of Electronic Administrative Data policy

<http://www.cit.cornell.edu/policy/drafts/InstData.html>

Cornell Spider

<http://www.cit.cornell.edu/security/tools/>

Return on Investment: While development and implementation costs will be inconsistent across Cornell (there are some colleges and units that already do an outstanding job of security) the investment will reduced cost associated with incident response, forensics and notifications as a result of computer compromises. Further, IT reliability and a reduction in lost time due to system rebuilds should be realized.

Replicable: 4

While many universities may not have the concept of the data steward and the associated coordination the concepts of the policy and the specific security requirements should be fully replicable.

Effectiveness: 3

We still need to tackle, in head-on fashion, the data access policy and procedures. While our current approach starts us down the right path it should not be considered a complete and thorough solution.

Notes: *Please enter any additional notes or comments.*

Contact for More Information: Steve Schuster, Director, IT Security Office, Cornell University, [sjs74@cornell.edu](mailto:sjs74@cornell.edu)