eTERA consulting

Is Personally Identifiable Information (PII) Pervasive on Your Company's Computers?

This white paper discusses:

- *What is PII and why you must manage it*
- *A case study of how one company addressed PII*
- *How you can begin to manage PII*

in this paper

The Authors

Mary Flory
Managing Director, Consulting Services

In her role as Senior Consultant, Ms. Flory is responsible for managing cases involving e-Discovery and large complex litigations. Prior to joining eTERA Consulting, Ms. Flory served as the Senior Manager of Practice Support at Dickstein Shapiro in Washington, DC. During her tenure at Dickstein Shapiro dating back to 1997, Ms. Flory specialized in large, complex MDL litigations, IP litigation and insurance litigation. She was a member of the firm's e-Discovery committee and managed the firm's litigation support operations and staff members. Ms. Flory has extensive experience assisting clients in establishing budgets and protocols for the collection of paper and electronic evidence. She consults with clients on how to plan, design and manage in-house and third-party computerized litigation support applications. Ms. Flory serves as the chair of eTERA Consulting's technology committee.

John Rubens
Managing Director, Corporate Analysis

In his role as Managing Director of Corporate Analysis, Mr. Rubens is responsible for assisting clients in taking proactive steps to manage the complex ESI landscape and effectively address their legal, compliance and regulatory requirements. With an extensive background in corporate analysis, including ESI indexing solutions, forensic analysis and acquisition, and Early Information Assessmentsm methodologies, Mr. Rubens brings a diverse background in legal, information technology and corporate risk management.

During his career, Mr. Rubens has successfully managed the completion of hundreds of data acquisition, forensic analysis, and litigation consulting projects around the globe, including such matters as patent enforcement, bid rigging, computer fraud, IP theft, embezzlement, counterfeiting, trademark infringement, electronic discovery and stock scams. He has extensive knowledge of large scale enterprise systems, software development processes and solutions, various operating systems and platforms, data backup and restoration software, and the use of forensic data acquisition and analysis technologies. Mr. Rubens is often involved in overseeing the creation of affidavits and support for court appearances related to the work of eTERA Consulting.

Prior to joining eTERA Consulting, Mr. Rubens was the Managing Director of The Oliver Group, an international forensic analysis company. He has also served as a Managing Partner at Senjiva, LLC.

One of the unintended consequences of the information age is the availability of Personally Identifiable Information (PII). The combination of name, date of birth, and social security number are the keys to the kingdom for the purposes of establishing false identity and fraud. Lost laptops, network break-ins, and phishing expeditions have led governmental entities to establish a patchwork quilt of laws requiring custodians of personal information to provide safeguards and assurance that PII is secure.

What is PII?

Personally identifiable information is any data about an individual that could, potentially identify that person, such as a name, fingerprints or other biometric data, email address, street address, telephone number or social security number.

A study done at MIT by Latanya Sweeney, now a professor at Carnegie Mellon University, found that 87% of the population in the United States could be uniquely identified by just 3 pieces of PII: their 5-digit zip code, gender and date of birth. This demonstrates that social security numbers, while valuable, is not necessary to identify unique individuals.

Now California agrees as well. On February 10, 2011, the California Supreme Court released its decision in *Pineda v. Williams-Sonoma Stores, Inc.*, holding that zip code information is personal identification information ("PII") under the Song-Beverly Credit Card Act (the "Song-Beverly Act"). The court's decision restricts businesses in California from requesting and recording a person's zip code as part of a credit card transaction.

Under the Song-Beverly Act, a business is prohibited from requesting, or requiring as a condition to accepting a credit card payment, the cardholder's personal information, which the business records. The California Court of Appeal had previously held that a zip code, without additional information, was not PII in *Party City Corp. v. Superior Court*. However, in Pineda, the California Supreme Court clarified California's broad interpretation of PII. An excellent legal recap of the California Supreme Court's ruling is available from Arent Fox.

Who Regulates PII?

Government Regulators

PII is coming at organizations from all angles – internal compliance requirements, federal and state requirements and, individuals suing over violation of their PII.

Regulatory agencies are concerned with PII. HIPPA was primarily created to protect patient PII. The Social Security agency has enacted regulations to protect Social Security numbers. States are enacting laws to protect data and the National Institute of Standards and Technology put out a Guide to Protecting the Confidentiality of PII in April of 2010.

Health Net had a data breach. They lost a hard drive with 27.7 million scanned pages of over 120 different types of documents such as insurance claims forms, membership forms, appeals and grievances, correspondence and

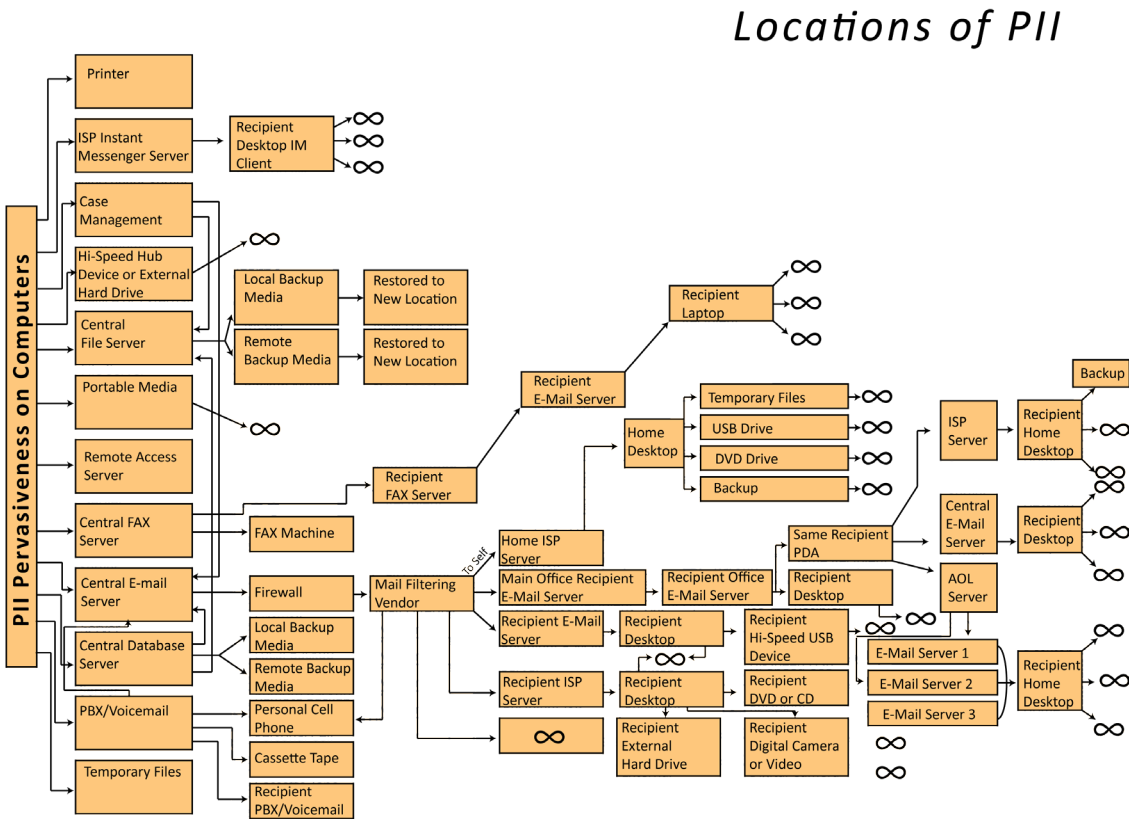
medical records. Health Net was fined $250,000, was ordered to put in an onerous plan of Correction in their offices by the Court. Additionally, Health Net spent more than $7 million to investigate the circumstances of the missing portable hard drive, to notify all the Health Net members of the breach and to offer credit monitoring services and identify theft insurance to anyone whose data has been breached.

Lawsuits have been filed for failure to protect PII. In *Claridge v. ROCKYOU*, ROCKYOU allegedly kept the email and passwords for its users in an unprotected plaintext file on the server which was hacked. Individuals sued ROCKYOU over failure to protect their PII.

The Need for a Data Topology Map

Breaches involving PII are hazardous to both individuals and organizations. Individual harms may include identity theft, embarrassment, or blackmail. Organizational harms may include a loss of public trust, legal liability, or remediation costs.

Fear of the unknown can keep you up at night – so it is best if you know what minefields your data might hold. While IT organizations excel at saving data, they are dependent upon other stakeholders, including legal and business units, to determine "what" should be saved. This process begins when the IT organization is able to communicate the various data types and locations where they reside.



As this diagram shows, there are lots of places where PII can reside within the organization. While many locations are areas the organization has control over – servers, printers, laptops, PII is often uncontrolled on employee home laptops, smart phones and USB drives.

A Case Study: Managing PII

In 2010, one of the largest utility companies in the U.S. completed an internal information audit at one of their sites and found some instances of employees storing personal information, particularly social security numbers, on email and file servers. The corporate IT organization selected eTERA Consulting to perform an analysis to determine the extent of the problem.

For this project, eTERA brought a leading information management system configured in a seven (7) server cluster that executed as a single system. In order to implement this system at the client location, all that was required was power, a single Ethernet cable connection and a temporary user account created with backup system rights on the network.

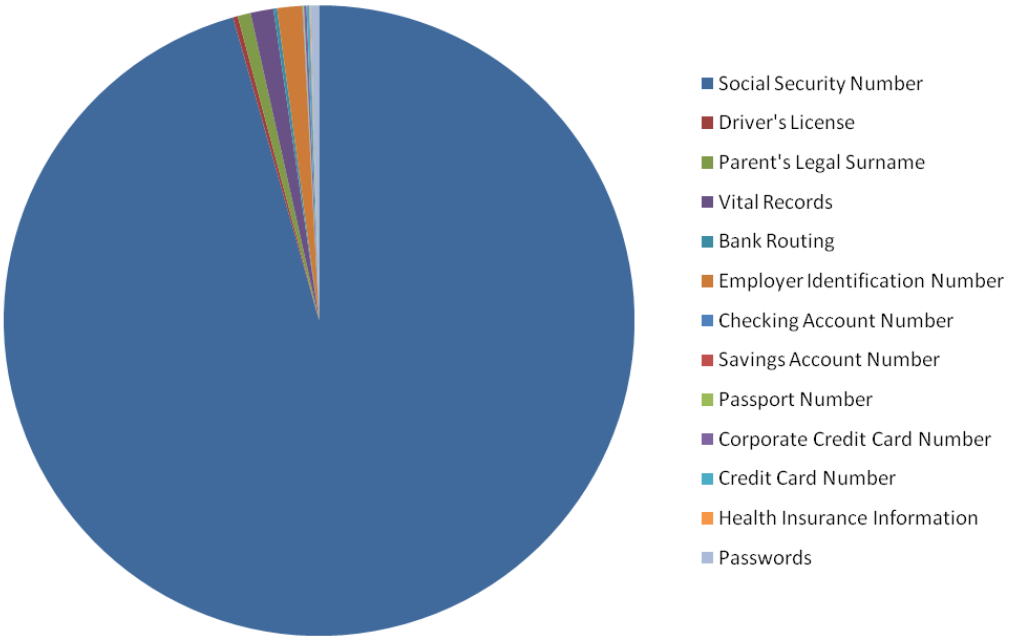
Once this was done, we preformed a quick crawl of all active servers on the network in order to identify locations and file types. Instead of fully indexing all of the files, the system created a thin index of file metadata resulting in a data topology map. This information allowed us to work with our client's IT group to specifically target areas of interest or ignore particular locations or file types that wouldn't apply to this project. For example, areas we excluded included a secured HR directory that contained employee records and any audio or video files that were on the servers.

Once our data targeting was complete, we completed full-text indexing. This process is scalable depending on the size of the servers and time frame of the project. For this particular project, our seven (7) server cluster indexed approximately 30 million items totaling about 17 TB in 5 days.

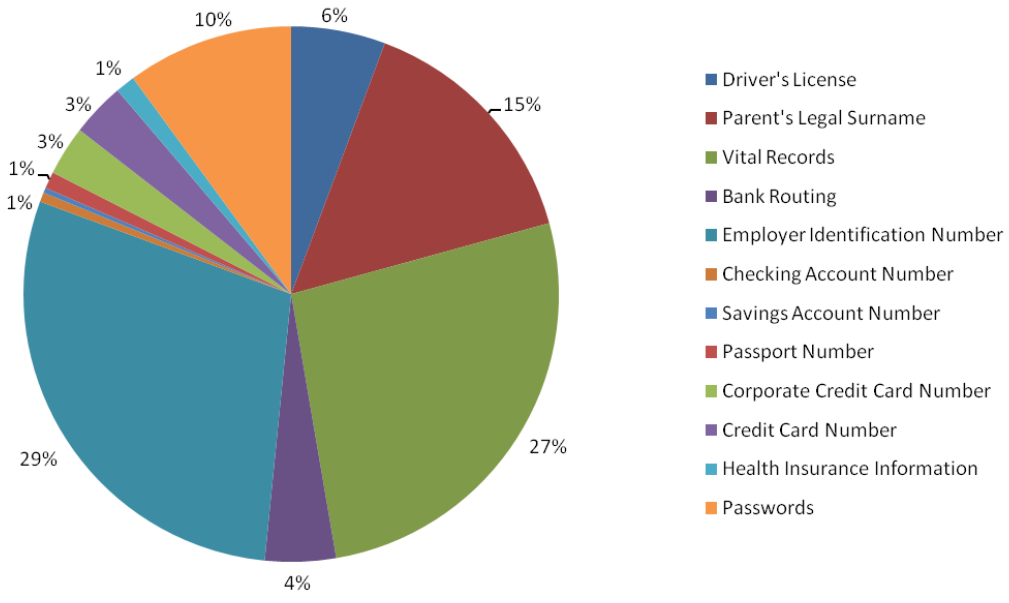
After the index was completed, we started running our queries. We started with a few predefined queries like social security numbers and vital records, which match words against a library of over 65,000 medical terms. We then wrote several custom queries to identify corporate credit card numbers, passport ID numbers, checking and savings account numbers, driver license number, employee ID's and passwords. Some of these required several revisions to improve their accuracy until they met with our client's satisfaction.

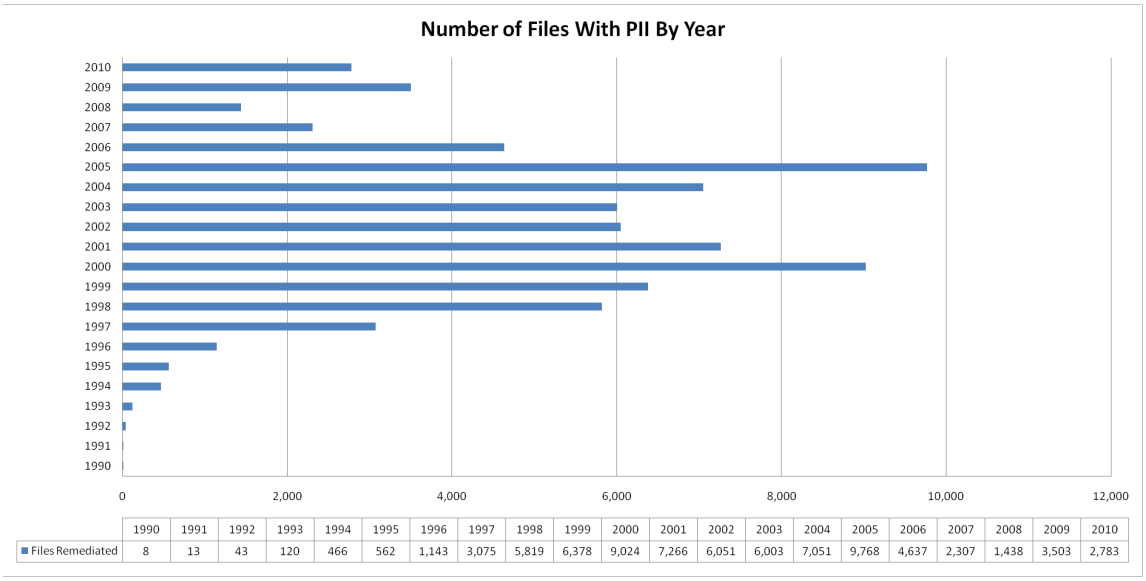
We then moved into the analysis of the results. Based on the internal audit our client ran previously, they only expected eTERA to find a few thousand files. What we found was far more than anticipated. A total of 76,100 files were found containing PII. This first graph shows that while there were more files with PII than originally estimated, these files were only a very small percentage of the total number of files on their servers and therefore a very manageable volume.

Of the 76,100 records that contained PII, 96% of those records contained Social Security Numbers.



When we take a closer look at the remaining 4%, another surprising result was the type of personal information that employees stored on the company's servers including personal banking information like checking and savings account numbers and even personal credit card information.





	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Files Remediated	8	13	43	120	466	562	1,143	3,075	5,819	6,378	9,024	7,266	6,051	6,003	7,051	9,768	4,637	2,307	1,438	3,503	2,783

Another interesting discovery was the age of some of these files. In this chart we organized the data by year and you can see that our client had records dating as far back as 1990 that contain PII.

PII Project Overview			
Total Number of Files Indexed on all 4 servers	**27,834,911**	**Total number of files identified from PII searches**	**76,100**

Files Remediated From PII Searches					
Data Element	**Server01**	**Server02**	**Server03**	**Server04**	**TOTAL**
Social Security Number	16,045	13,020	25,922	17,334	72,321
Driver's License	34	70	87	53	244
Parent's Legal Surname	88	132	112	211	543
Vital Records	87	467	190	184	928
Bank Routing	24	37	34	78	173
Employer Identification Number	139	258	616	54	1,067
Checking Account Number	12	7	13	6	38
Savings Account Number	6	4	4	6	20
Passport Number	3	12	22	9	46
Corporate Credit Card Number	19	8	9	94	130
Credit Card Number	41	50	9	33	133
Health Insurance Information	21	23	4	23	71
Passwords	74	87	180	45	386

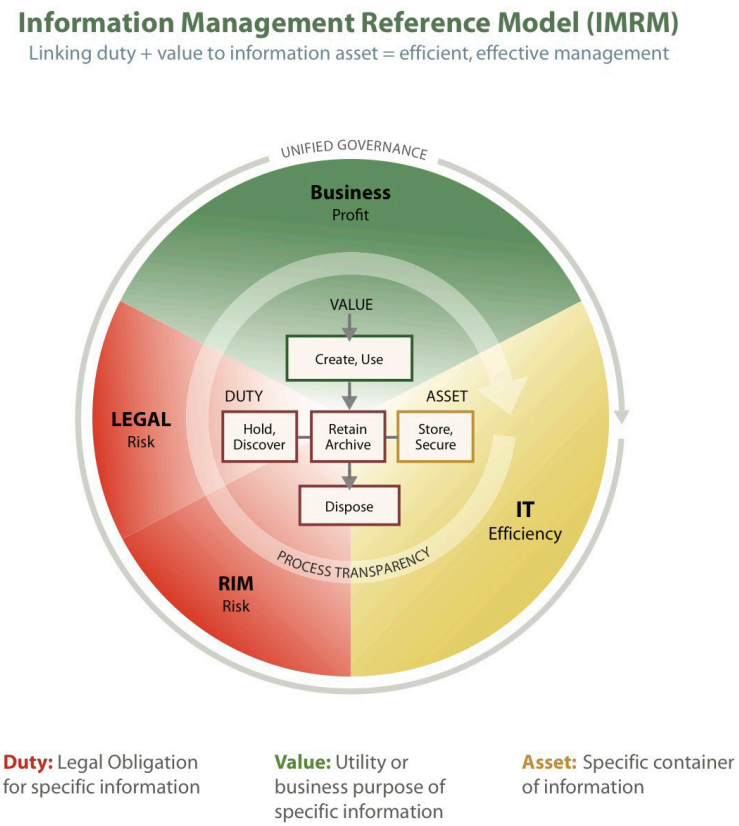
The final step in this project was remediation. For this project, our client decided to preserve the data before deleting. We copied all of the targeted records containing PII to a secure location while maintaining the directory structure. Once that effort was completed, we deleted the records from their original locations. We also provided a defensible audit report for every phase of this project.

PII and Information Management

It is critical to remember that having PII is not really the issue. It is how it is managed and controlled that determines whether or not you have a problem.

PII is also a fairly narrow topic, and it really needs to be considered in the context of your overall information management and governance program. Most of you are familiar with the Electronic Discovery Reference Model (EDRM). There is more recent initiative, the Information Management Reference Model (IMRM) depicted here that focuses attention on the far left of the EDRM model.

The IMRM model considers the key information management stakeholders in an organization – Business, Information Technology, Risk Management and Legal – and the context in which each has a responsibility in managing



and governing organizational data assets. There are Business requirements for how information is created and used and the value it brings; there are technical considerations associated with how data is managed, stored and secured; and there are Risk Management and Legal considerations for how data is used, or misused. All of which has a cost associated with it. Certainly the cost for not managing it well is significantly higher.

In addition, there a many policies, procedures and processes that may already be taking place that can be useful in looking further at the PII risk. There is often cross-over with other information governance and security measures such as Data Loss Prevention and Intellectual Property Management, or Human Resources programs and controls.

The key take away from this is that your initial focus may be PII, but that there may already be things in place that you can leverage – although if there aren't, then you should also make sure you at least consider the bigger picture and include the key stakeholders when embarking on a PII information governance initiative.

Getting Started

You've now read the results from one of our PII projects. Now let's walk through the high level approach we took and also provide you with some key points to consider in getting started.

Our engagement began with an analysis with our unique Early Information Assessment℠ (EIA) methodology. EIA is a methodology that integrates people, planning, resources and innovation to help legal, IT and business stakeholders improve the overall management of information management and E-Discovery operations while achieving cost-savings. Two of the key EIA program components include:

7/12 Analysis – eTERA's EIA team evaluates "7" basic corporate requirements and integrates a combination of "12" critical innovative solutions to achieve desired outcomes.

BOE Analysis – The EIA team assesses the client's "BOE" or Business Operating Environment. The BOE encompasses technology, staffing, training and education, project management, resources, tools, workflow, processes, reporting, communications, stakeholders, vendors, security and budget. A total legal department management assessment is conducted by the EIA team.

If you are doing this yourself, you really need to define the problem – for example, is this a proactive or reactive effort? Are you looking for PII because you know it exists where it shouldn't or was there an event that is triggering the evaluation? As you read in our example, the initial problem definition was only partially correct and understood. That's OK, because in going through the process the potential issues and risks were further refined and adjusted. But again, in looking at any potential PII issues, it is always good to bring in the key experts and stakeholders who can assist – if those people are available internally and know the right questions to ask, then that's

The Approach

Define the problem

- Use knowledgeable experts in technology, legal and regulatory

Define the scope

- Narrow target / Short term
- Broad search / Long term

Leverage existing libraries and known data structures

- SSNs, credit cards, healthcare info, etc.

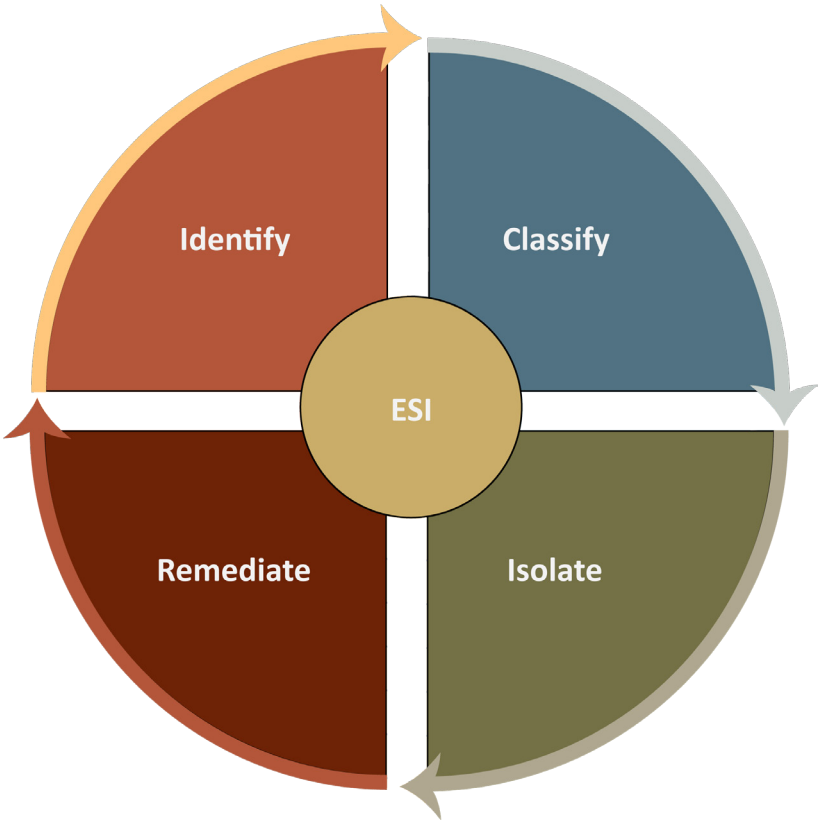
great. If you don't have the expertise in-house, then you can look outside your organization for people that can assist – as was the case with our client.

Next, you need to think about how Narrow or Broad the effort needs to be and whether or not this a one-time, short-term initiative; or is it a long-term process that will be integrated into your overall information governance and control structure?

You could start simply with a single server that you've identified; or there may be specific a specific Business Unit that is an area that has the highest likelihood of harboring PII; or you could decide there is enough risk to warrant an Enterprise-wide search. This decision has an impact on how you will end up moving forward, as well as the best technology to use, and ultimately the costs.

Lastly, in an effort to make the process as efficient as possible, you need to leverage existing information and criteria. As in our example, that could be existing libraries or known data structures of PII; or developing those specific examples that may be unique to your organization.

When we look at a PII initiative in this context we like to work through the following process:



Identify

We start by using sampling methodologies to quickly rule-in or rule-out the need for further indexing and testing. We also use pre-formulated queries and libraries that speed the process.

Classify

Once we've identified potential PII, we need to Classify the data. In many instances, this can be automated through the use of technology and the libraries. In other instances it may require human intervention (e.g. a review process) to further classify the data as being PII or not.

Isolation

If we do identify PII, we look to Isolate that data. In some cases this is done in place, where the data is currently stored. In other cases it involves moving of the data to another location for further analysis.

Remediation

The last stage is Remediation. During this stage we consider the ultimate disposition, or what we want to do with PII that we've identified. This could entail implementation of additional security measures to prevent unauthorized access. Or it could involve securely deleting the data…or any combination the above.

Conclusion

In looking at the ultimate benefits achieved by the client using the EIA approach and a Managed Services model, here's how they were able to measure the value:

- First, and most importantly, they found the PII they were looking for and were able to successfully remediate it, eliminating their risk and potential issues in the future.

- In terms of human resources and capital, they ultimately got the value of having 10 experts for the cost of 2. Since the experts included various technical, risk management and legal disciplines, the value was even greater.

- From the Process standpoint, the problem and potential solutions were quickly identified. We then went through an iterative process to fine tune the solution and achieve the results. We also provided a sort of "Neighborhood Watch" in monitoring things as they progressed and were able to intervene quickly as needed.

- And from the Technology standpoint, we were able to offer several different options from several different solution providers and then recommend the best solution based on performance and cost benefits. This provided more options to the client and lower risk as they didn't have to make a long-term decision to get things started.

About eTERA Consulting

eTERA Consulting specializes in helping organizations improve information governance, compliance and discovery management. We are a technology independent consultancy with a broad range of services from strategic information consulting to project-based engagements. Because we are not locked into a particular vendor's technology, our clients benefit from flexible service delivery and pricing options.

A unique differentiator for eTERA Consulting is our Early Information Assessment℠ (EIA) methodology which we use to help clients implement a proactive approach to integrating the management of corporate information as it relates to risk management, regulatory compliance, electronic discovery, litigation hold, records management and IT storage.

Headquartered in Washington, DC, eTERA has served the legal industry since 2004. The company also maintains additional offices across the United States. eTERA was named to The Inc. 500 list of fastest-growing private U.S. companies in 2010.

DISCLAIMER: Legal Information Is Not Legal Advice

eTERA Consulting is not a law firm. This bulletin provides information to help professionals better understand the intersection between technology and legal processes. But legal information is not the same as legal advice -the application of law to an individual's specific circumstances. Although we go to great lengths to make sure our information is accurate and useful, we recommend you consult a lawyer if you want professional assurance that our information, and your interpretation of it, is appropriate to your particular situation.

eTERA Consulting
1100 17th Street, NW
Suite 605
Washington, DC 20036
Direct: 315-566-9330
info@eteraconsulting.com
www.eteraconsulting.com