Only a Matter of Time?

Incident Response & Defensible Data Breach Discovery

Educational Webcast

02 | 17 | 2021

HAYSTACK

Michael Sarlo

Chief Innovation Officer & President of Global Investigation Services for HaystackID

Michael facilitates all operations related to electronic discovery, digital forensics, and litigation strategy both in United States and abroad while working on highly complex forensic and e-Discovery projects. He has full oversight of all facilities and manages workflow and change management to ensure consistent quality and efficiency of all processes for each project entering HaystackID's walls.

John Brewer

Chief Data Scientist for HaystackID

John Brewer has been a software engineer and information technology worker for over 20 years and worked for dozens of Fortune 500 firms in roles from eDiscovery to Data Migration to Information Stewardship.

He's worked with Haystack ID since 2015 on bringing the latest advancements in internet technologies to the eDiscovery and IR market.

John Wilson

CISCO & President of Forensics for HaystackID

John provides expertise and expert witness services to help companies address various matters related to digital forensics and electronic discovery (eDiscovery), including leading investigations, ensuring proper preservation of evidence items and chain of custody. He develops processes, creates workflows, leads implementation projects as well as GDPR data mapping services for clients including major financial institutions, Fortune 100 companies, AmLaw 100 law firms as well as many other organizations small and large. In addition, he provides expert witness services and consulting in matters of all sizes. His work spans some of the largest litigations and matters on record in the United States and many of the 39 countries where has worked on cases.

Jennifer Hamilton

Deputy General Counsel for Global Discovery & Privacy for HaystackID

Jennifer serves as a resource for corporate clients, support legal and compliance operations, and continue to grow the Enterprise Managed Solutions Group, the company's specialized offerings for corporations and law firms wishing to transform their business of law practices. Jennifer comes from John Deere, where she spent 14 years leading the development of the company's eDiscovery operations and was head of the Global Evidence Team.

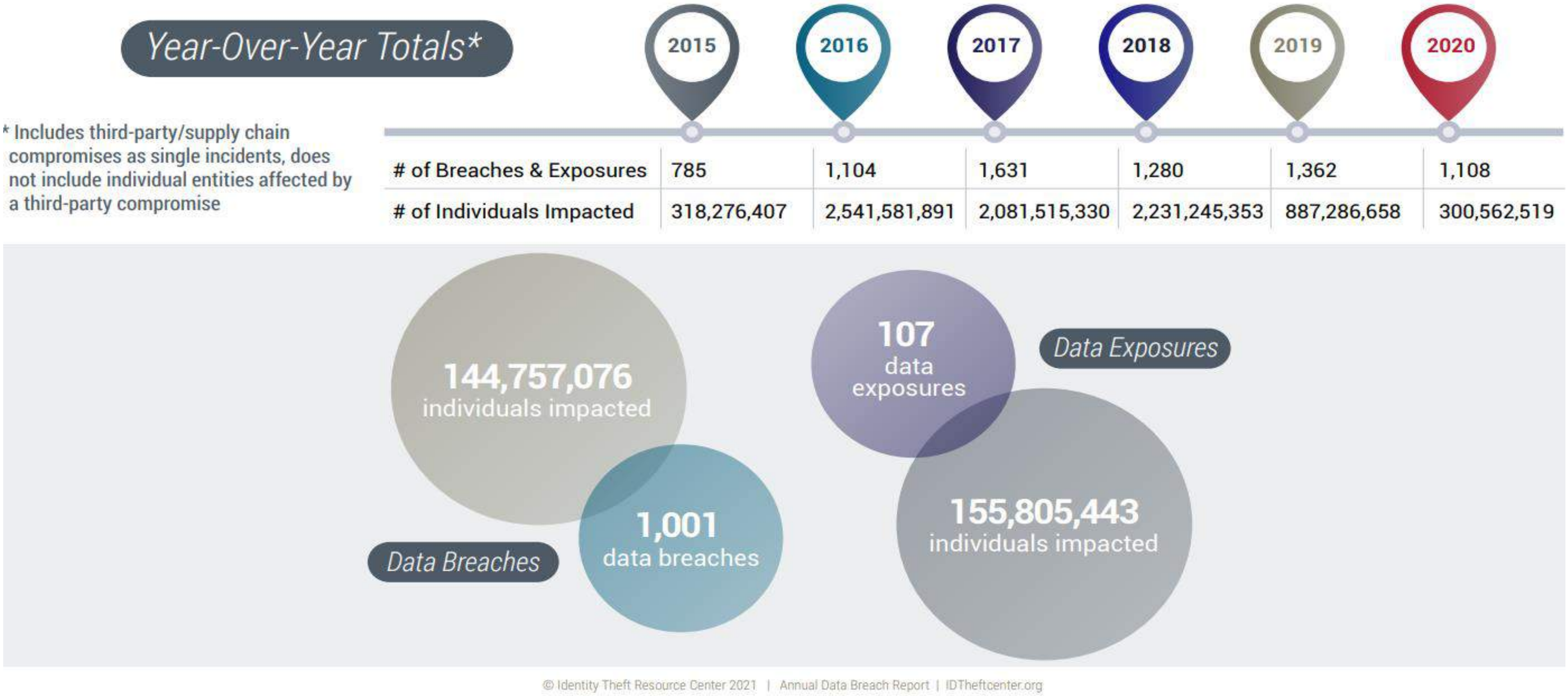
Agenda

- It’s Only A Matter of Time: Security Incident Statistics & Ransomware
- The First 48 Hours: Electronic Security Incident Detection & Incident Response
- Effective IR Plan Design: Simplicity, Scalability & Moving Beyond the Breach
- Post-Breach Discovery: Workflow, AI, and Impact Assessment
- ReviewRight Protect: Post Breach Review & Entity Extraction Workflow

It’s Only A Matter of Time:

Security Incident Statistics & Ransomware

Number of Compromises



*Year-Over-Year Totals**

	2015	2016	2017	2018	2019	2020
# of Breaches & Exposures	785	1,104	1,631	1,280	1,362	1,108
# of Individuals Impacted	318,276,407	2,541,581,891	2,081,515,330	2,231,245,353	887,286,658	300,562,519

* Includes third-party/supply chain compromises as single incidents, does not include individual entities affected by a third-party compromise

Data Breaches

144,757,076
individuals impacted

1,001
data breaches

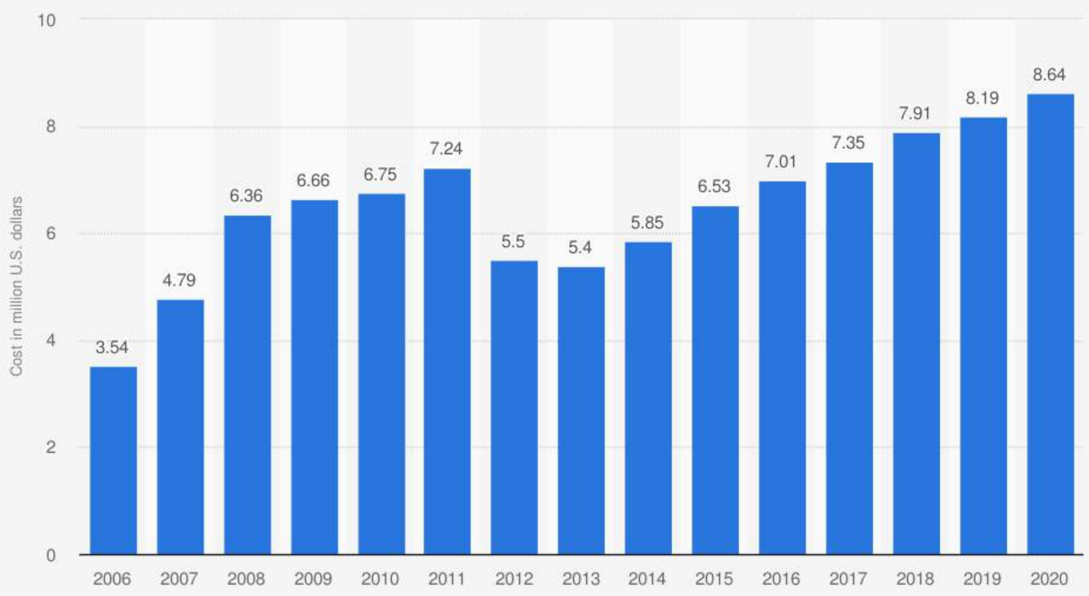
Data Exposures

107
data exposures

155,805,443
individuals impacted

© Identity Theft Resource Center 2021 | Annual Data Breach Report | IDTheftcenter.org

Average Cost of A Breach



Average organizational cost to a business in the United States after a data breach from 2006 to 2020 (in million U.S. dollars)

Sources
Ponemon Institute; IBM; ESET North America (Welivesecurity.com)
© Statista 2021

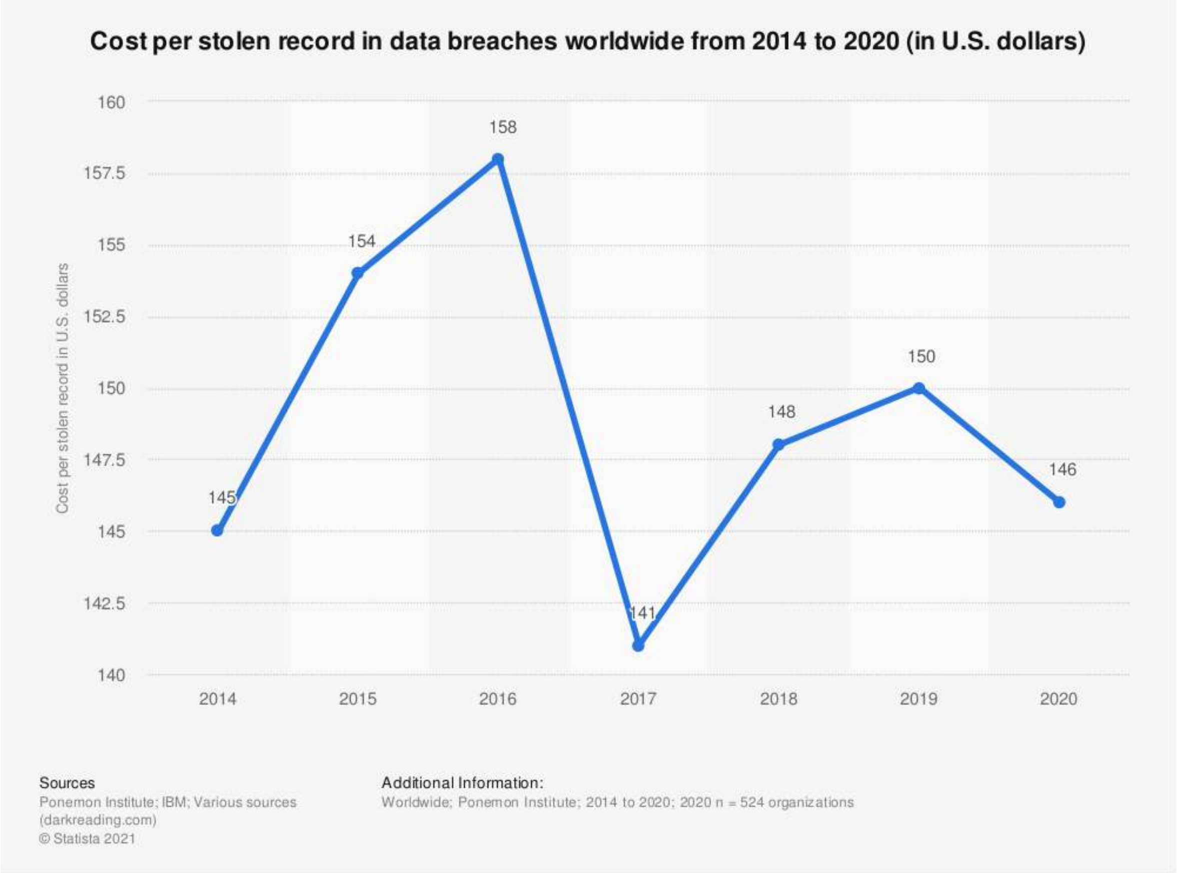
Additional Information:
United States; Ponemon Institute; IBM; 2006 to 2020

In 2020, the average cost to businesses affected by a data breach in the United States amounted to 8.64 million U.S. dollars, up from 8.19 million U.S. dollars in the previous year.

The global average cost per data breach was 3.86 million U.S. dollars.

Total breach costs include: lost business resulting from diminished trust or confidence of customers; costs related to detection, escalation, and notification of the breach; and ex-post response activities, such as credit report monitoring. The date is dated in the year of publication rather than the fieldwork completion date.

Average Cost per Record



Sources
Ponemon Institute; IBM; Various sources (darkreading.com)
© Statista 2021

Additional Information:
Worldwide; Ponemon Institute; 2014 to 2020; 2020 n = 524 organizations

In 2020, the cost per stolen record in data breaches was amounted to 146 U.S. dollars, down from the all-time high of 158 U.S. dollars per stolen record in 2016.

The sector with the highest cost per stolen record in data breaches was healthcare, which had a cost of 429 U.S. dollars per stolen record due to a data breach.

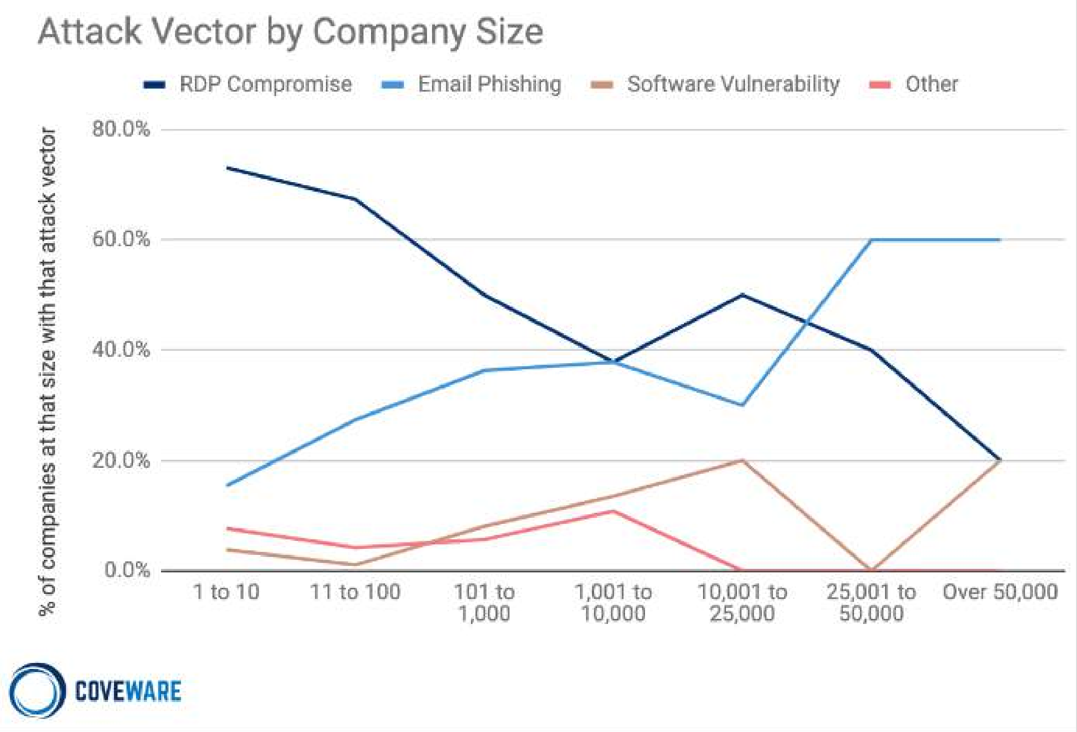
The Anatomy of a Ransomware Attack

Ransomware is essentially a virus that loads onto a user's computer, where it scans connected drives for files that it then encrypts. The user is also typically locked out of their machine and can only view a screen showing how to make a ransom payment.

Ransomware attacks can take many forms, although the most common is to prevent a user from accessing encrypted files or using their machine until the ransom is paid (cryptocurrencies preferred). More malicious ransomware attacks threaten to release sensitive data to the internet broadly (doxware) or to delete data permanently.

Ransomware can reach a user's machine using a number of vectors, the most common of which is a phishing attack. However, malicious websites or popups may also provide access for ransomware attacks. Ransomware attacks can also be directly injected into an organization's network through unsecured network connections (i.e. if no VPN is used). Or, even more simply, criminals may simply use brute force to hack weak passwords and directly insert the ransomware themselves.

Ransomware can also attack vulnerabilities in applications arising during the software development process. It is therefore important to use testing methods, such as static and dynamic application security testing (SAST/DAST), that identify these security vulnerabilities continuously while your applications are running.



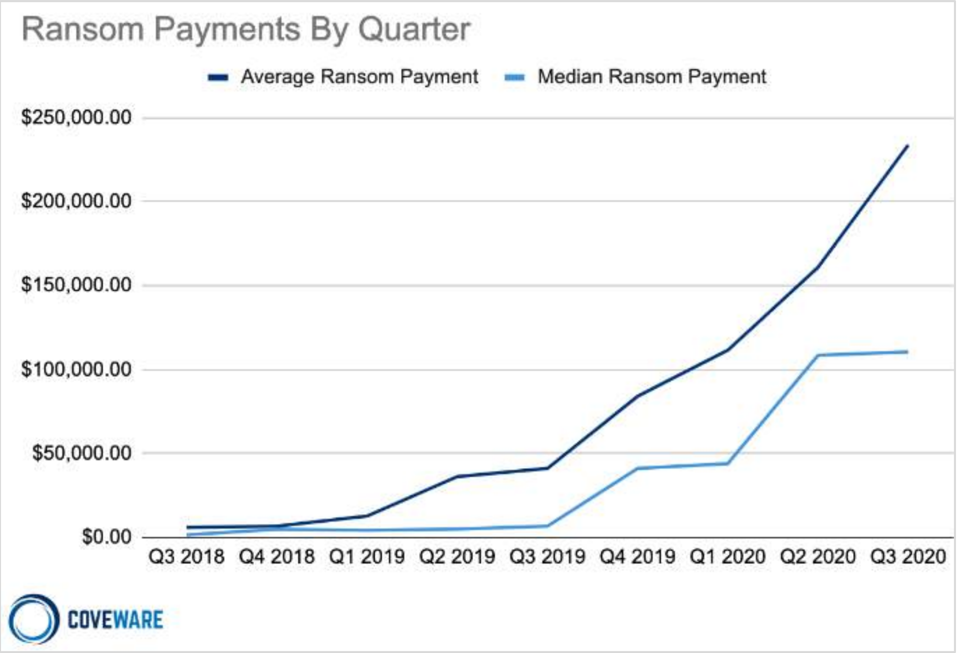
Average Ransom Payment Sizing

Analysis of 2020 data breaches reveals the continuation of a trend from 2019: cybercriminals are less interested in stealing mass amounts of consumers' personal information. Instead, threat actors are more interested in taking advantage of bad consumer behaviors to attack businesses using stolen credentials such as logins and passwords.

Ransomware and phishing attacks directed at organizations are now the preferred method of data theft by cyberthieves. These attacks generally require only a stolen credential or for an employee to click on a link in an unsolicited email, text, or social media account. Ransomware and phishing require less effort, are largely automated, and generate payouts that are much higher than taking over the accounts of individuals. One ransomware attack can generate as much revenue in minutes as hundreds of individual identity theft attempts over months or years

The average ransomware payout was > $233,000 per event in Q4 2020.

- Average Ransom Payment $233,817 +31% from Q2 2020
- Median Ransom Payment $110,532 +2% from Q2 2020



The First 48

Electronic Security Incident Detection & Classification

The First 48: Ransomware

Signs you are about to get hit

1. **Partial MFA Logins** – Passing Password but not 2FA

2. **Brute-Force Attacks** Will Hit the Network

3. **Phishing Emails** Land with Strange Domains

4**. Jump Boxes** Start Spinning Up

5. **SMB, Kerberos,** or **LDAP** Requests Come from Unexpected Appliances

6. **Broadcast Traffic** from P2S VPN Connection

7. Abrupt Increases in **Non-HTTPs Outbound Traffic** From Client Machines

The First 48: Ransomware

You’ve been hit, what’s next? First Calls

You’ve confirmed data has been accessed…

You’re not sure how much has gotten out…

You have no idea what your legal exposure or responsibilities are…

Information Technology – **STOP THE LEAK**

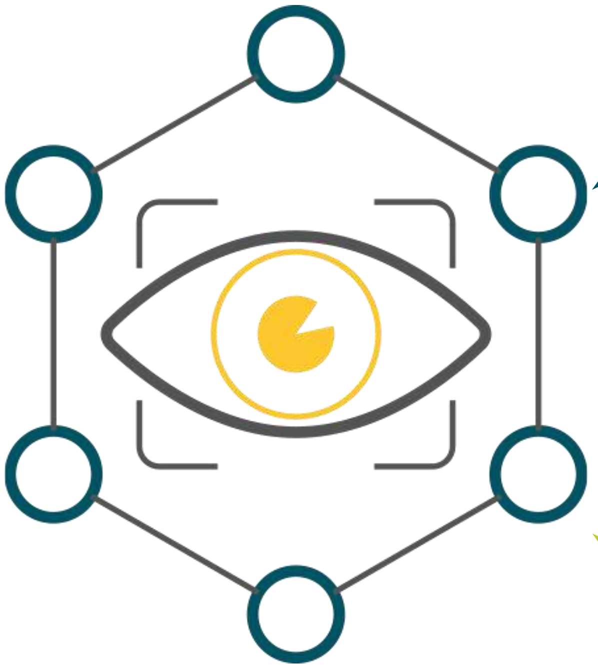
- **Don't wait!** Use that emergency line, use a personal number, ignore vacations, ignore sick days, get in contact with the person who can seal the leak.
- **Change the password** on whatever account might have been compromised.
- **Halt all systems** that automatically rotate or delete old logs and preserve of everything you can.
- **Secure all backups** – start moving off-sites back to the site.

The First 48: Ransomware

You’ve been hit, what’s next? First Calls

WHO was exposed?
Was our customer data taken?
Was our employee data taken?
Was our vendor data taken?
Are there other people’s info we’re responsible for?

WHEN did the attacker get in?
What time were they locked out again?



WHAT did we do?
Document everything you do in response to the attack, especially in the first few hours and days.
Almost any good-faith action you take will help you later.
DO NOT delete anything that isn’t an immediate threat.

WAS anything altered?
Was data changed?
Was anything installed?
Were any new accounts created?
Were permissions changed?
Is there a persistent threat?

WHAT did they have access to?
What’s the best-case scenario?
The worst-case scenario?

Effective IR Plan Design

Simplicity, Scalability & Beyond the Breach

Key Players

Establish a Scalable Bench of In-House & External Resources

- Internal IT lead
- Internal legal lead
- Outside legal counsel
- Outside forensics firm
- Other?

Define When Outside Counsel Needs to Run Point

List Key Roles (not just names) on the Plan

Define Who is on the Core Team Versus Expanded Team

Key Workstreams

Draft Concise Workflows for Highest Risk Events

- Assembling team to pinpoint breach notification requirement & recommendations
- Engaging insurance & legal
- Communication with the board

Indicate Which Workstreams can be Run in Parallel

Communication Plan

Craft a 1-Page Plan By Role

The Fewer Players, the Better

- Pre-vet outside counsel with insurer
- Have an MSA with multiple global vendors (important when breaches scale)

Leverage Crisis/Communications Teams

Data Mapping

Previous Data Mapping Can Support Incident Response

- Tagging in O365 and record retention schedules
- GDPR/CCPA data mapping
- Institutional eDiscovery knowledge

Remediation Efforts Should Flow From Most Sensitive Data Repositories

Response & Notification

Parallelize Workflows & Compartmentalize Risk Along Potential Notification Requirements

- Geography & Data Types

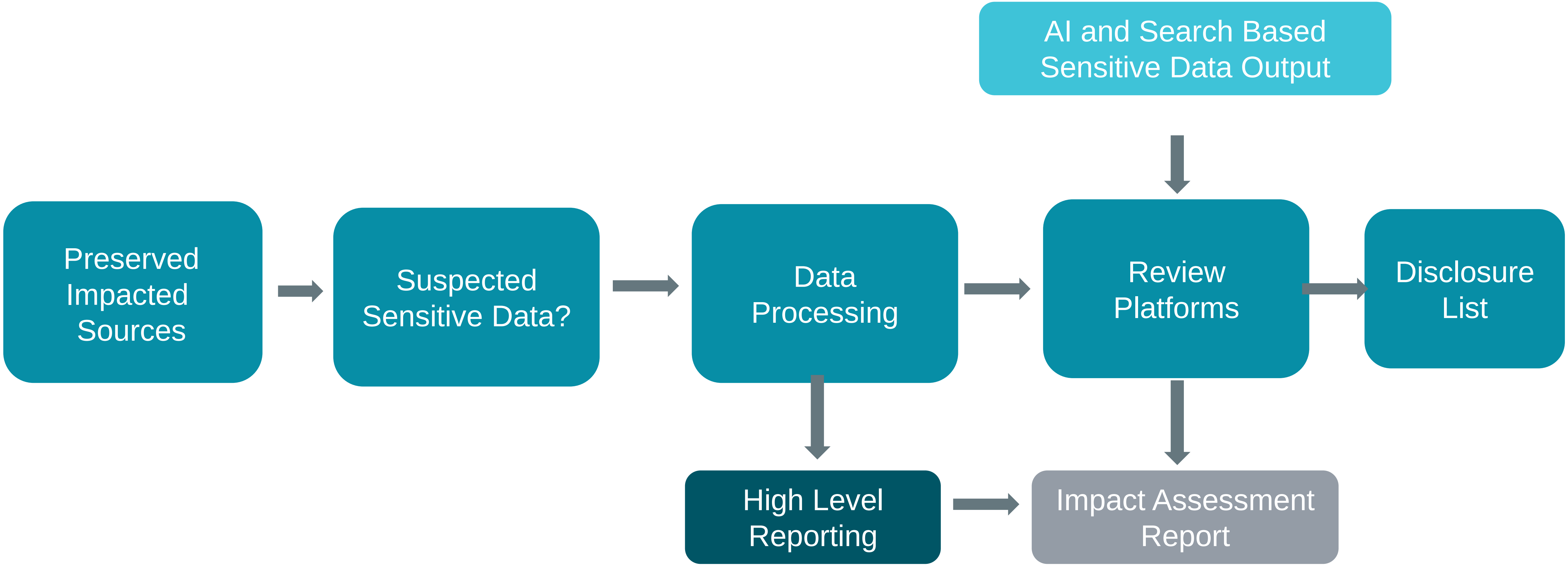
To Speed Up Response Around Notification Requirements, Engage One Stop Shops

- IR, eDiscovery, digital forensics and document review

Post-Breach Discovery

Workflow, AI & Impact Assessment

Post-Breach Discovery Workflow



AI and Search Based Sensitive Data Output

Preserved Impacted Sources

Suspected Sensitive Data?

Data Processing

Review Platforms

Disclosure List

High Level Reporting

Impact Assessment Report

Multi-Engine Entity, PII, & PHI Detection & Extraction

Modern Techniques

- Word2Vec
- Template Matching

Cutting Edge

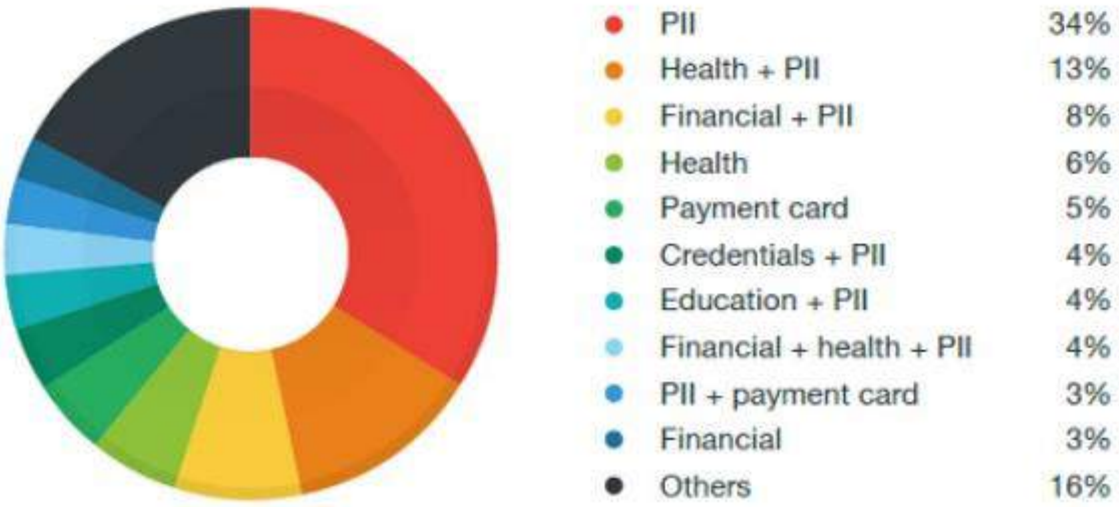
- Augmented Transcripts
- GPT Models
- Triumvirate Cognitive Models
- Sentiment Detection
- Non-Entity Key Phrases

Stock Sensitive Data Breach Assessment Reporting

Automated Customizable Impact Assessment Reporting

AI Engines and Search Workflows Allow for Creation of Customized Reporting that Includes:

- Count of Sensitive Data by Type
- Count of Sensitive Data by Source
- Count of Unique Person Names and Organizations
- Count of Unique Persona Names and Organizations that Overlap with Sensitive Data Types
- Count of Sensitive Data by Range of Confidence Scores
- Count of Document Types within the Above Categories
- Count of Sensitive Data by Type over Custom Date Ranges
- Roll up reporting of Top Folder Locations
- General Dataset Statistics
- Visual Reporting via Customizable Dashboards
- Exception Reporting
- Deduplication Statistics



ReviewRight Protect

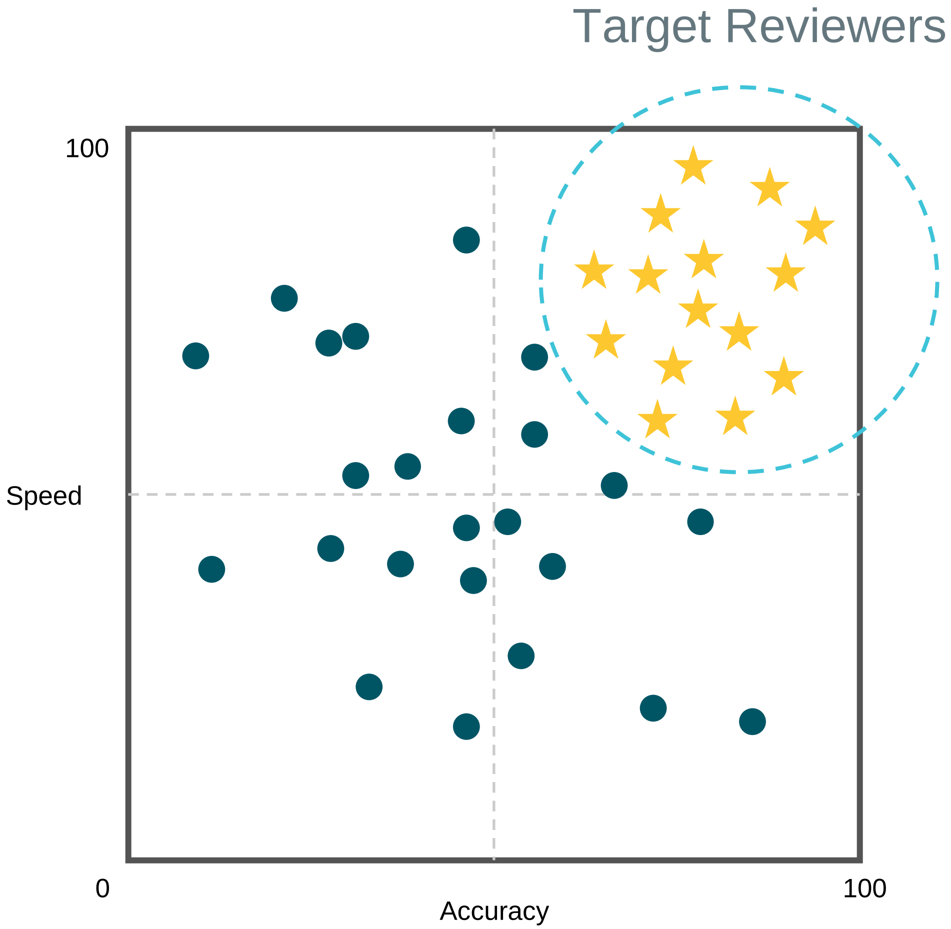
Post-Breach Review & Extraction Workflow

Reviewer Selection

Qualification

ReviewRight Test Assessment

Reviewers who seek to be considered for document review opportunities must first take a skills assessment test. They are presented with a fact pattern (i.e. case background) and a review protocol. With this information, the reviewers are **administered 15 documents and are asked four (4) questions** related to relevance, issue spotting and privilege per document that provides HaystackID with **immediate, quantifiable, objective insight into a reviewer's legal review capabilities, including projected accuracy per task and rate of review**.



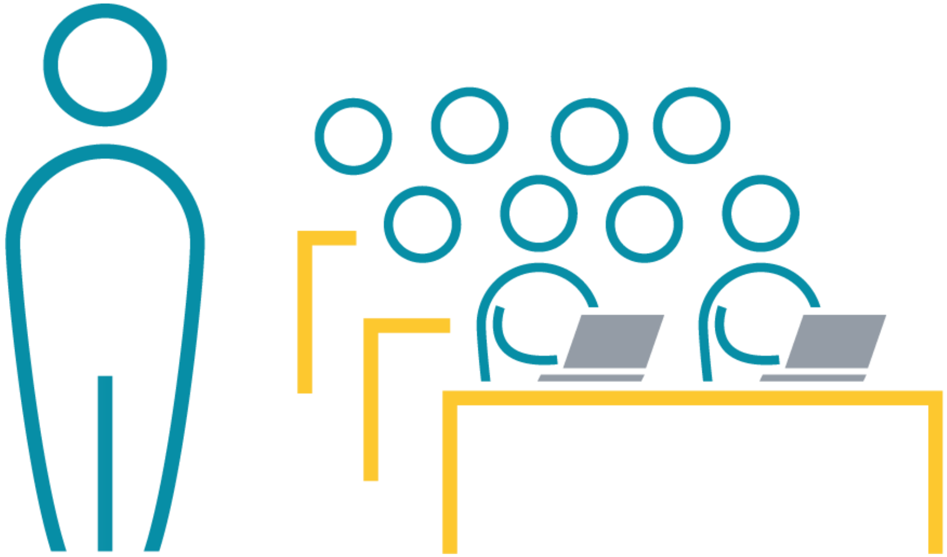
HAYSTACK

Quality Assurance

Starting with the Gauge Analysis

Gauge Analysis

Before releasing team members to pull general review batches, we use a gauge analysis that **tests and scores each reviewer's coding on the same set of documents**. Using this test ensures that the reviewers understand the review protocol before being released to code the review set. In addition, we use this test to **identify protocol deficiencies** by gauging the reviewers' responses against counsels' responses. Allowing us to get immediate feedback for the team and adjust the protocol if necessary.

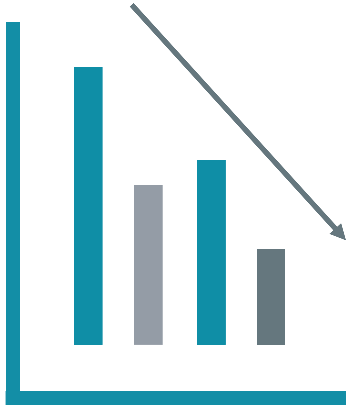


HAYSTACK

Breach Review Philosophy

ReviewRight

HaystackID Review Managers work closely with counsel, clients and data experts from the early stages of a breach through final disclosure reporting to devise document review and data extraction strategies that **limit review set populations,** increase review speed and accuracy, while **reducing false positives,** via **customized workflows** that leverage **human expertise** that are **enhanced via machine learning** and **advanced data analyses techniques.**



Reduced Review Counts

Workflows

HAYSTACK

Baseline Review Set Reduction & Optimization

Towards a Smaller Set Needing Human Review

Saving Time and Money

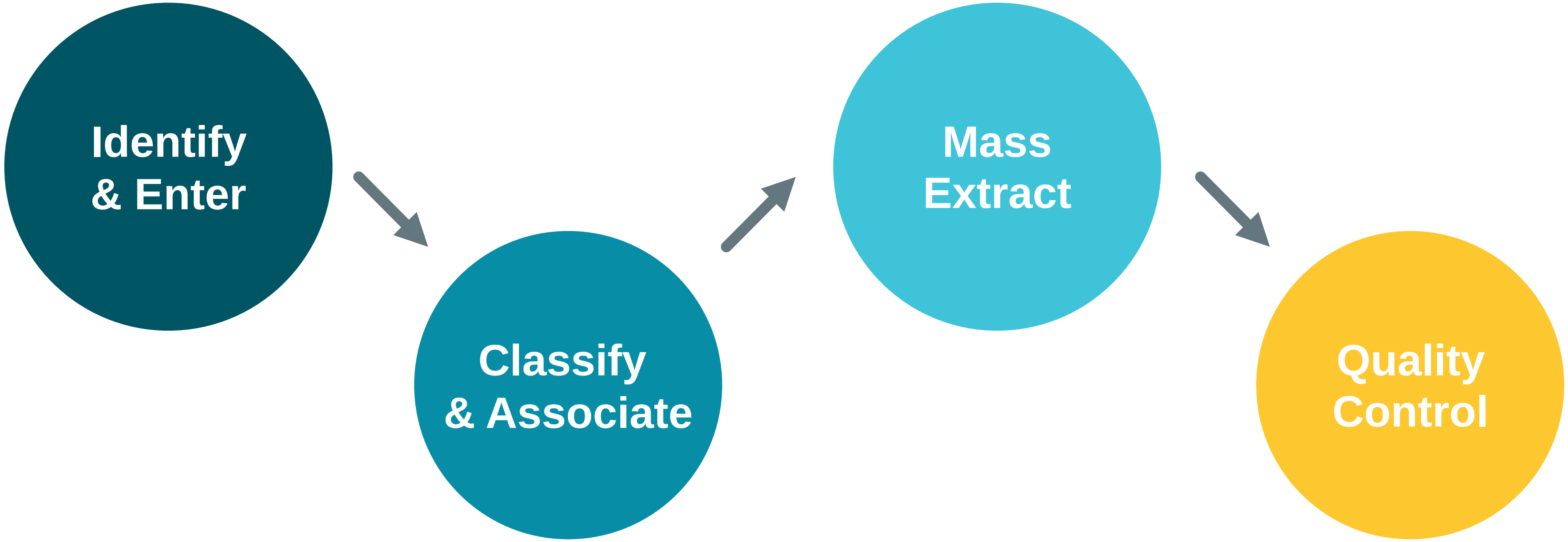
Any reduction in data size leads to immense cost and time savings, as human review far outstrips machine team in baseline cost. HaystackID **aggressively pursues data reduction** through a variety of means, tailored for your data, including:

- *Deduplication/Suppression of Documents*
- *Domain analysis*
- *Repeating form exclusion*
- *Search term analysis*
- *Regular expression/pattern searches*
- *Batching for speed and accuracy*

All optimized with inputs from machine learning analysis

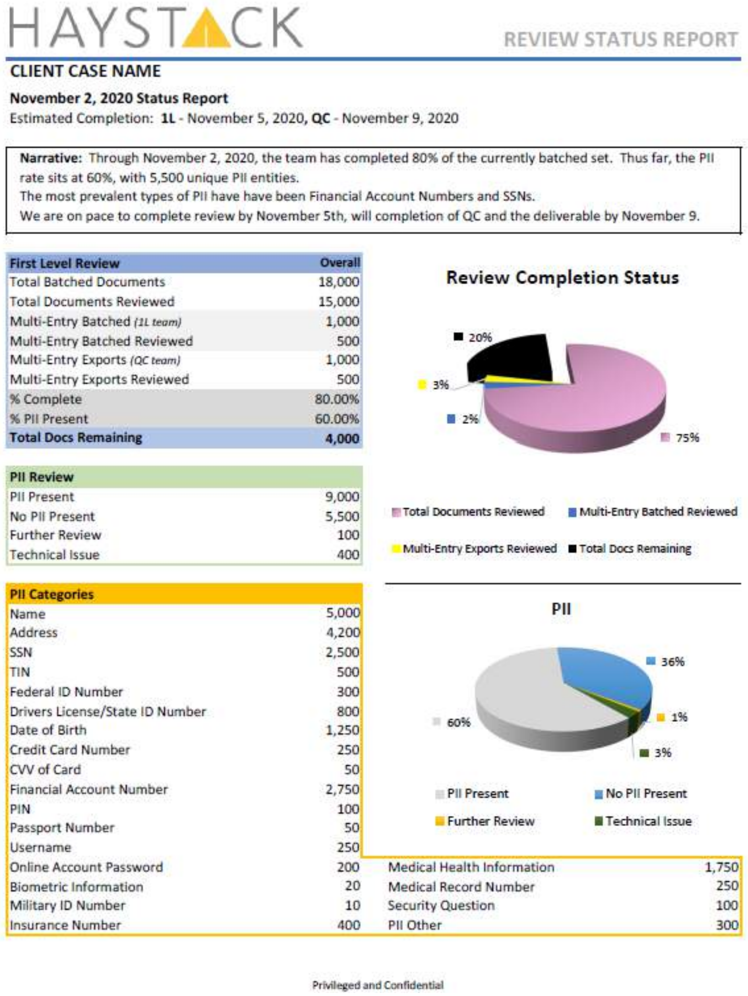
Review Methodology

Individual and en masse



Reporting

Along with an up-to-date issue log, every day, HaystackID provides clients/counsel with customized project review metrics for all coding fields and choices, unique entity counts, estimated completion dates, QC metrics, individual and team pace and overturn rates, as well as a detailed narrative that provides key information as to the status of the review.



HAYSTACK REVIEW STATUS REPORT

CLIENT CASE NAME

November 2, 2020 Status Report
Estimated Completion: **1L** - November 5, 2020, **QC** - November 9, 2020

Narrative: Through November 2, 2020, the team has completed 80% of the currently batched set. Thus far, the PII rate sits at 60%, with 5,500 unique PII entities.
The most prevalent types of PII have have been Financial Account Numbers and SSNs.
We are on pace to complete review by November 5th, will completion of QC and the deliverable by November 9.

First Level Review	Overall
Total Batched Documents	18,000
Total Documents Reviewed	15,000
Multi-Entry Batched *(1L team)*	1,000
Multi-Entry Batched Reviewed	500
Multi-Entry Exports *(QC team)*	1,000
Multi-Entry Exports Reviewed	500
% Complete	80.00%
% PII Present	60.00%
Total Docs Remaining	**4,000**

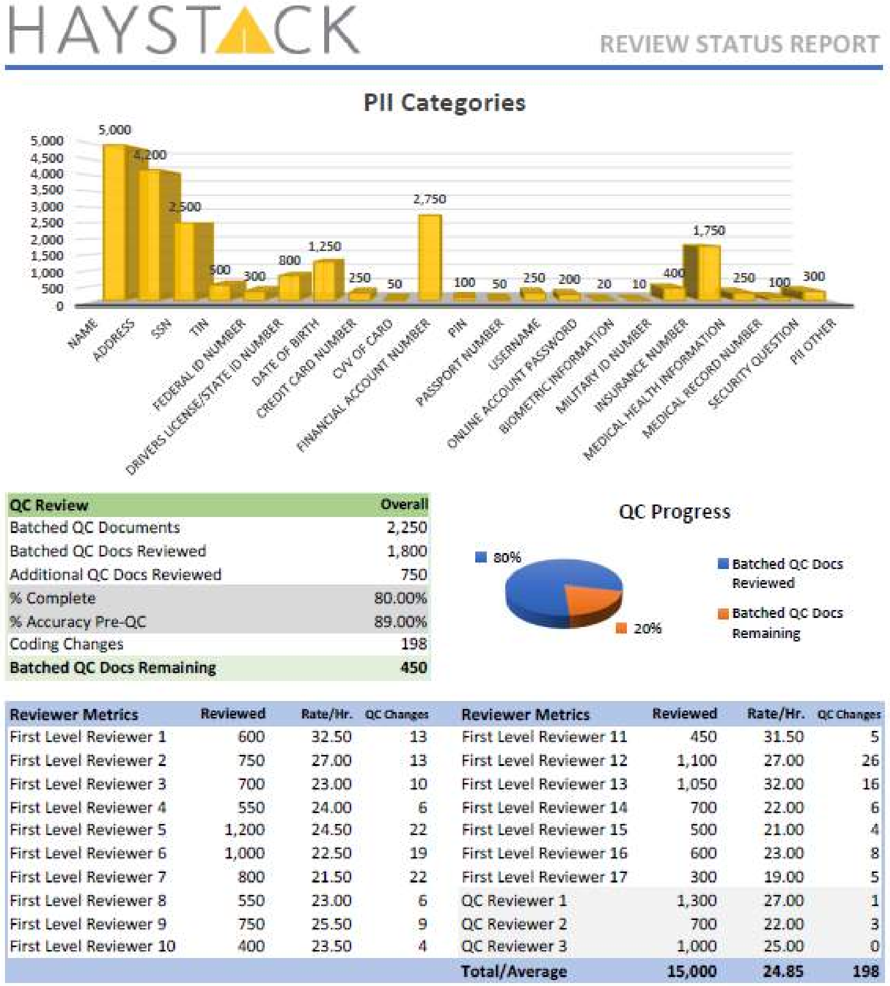
Review Completion Status

PII Review	
PII Present	9,000
No PII Present	5,500
Further Review	100
Technical Issue	400

PII

PII Categories	
Name	5,000
Address	4,200
SSN	2,500
TIN	500
Federal ID Number	300
Drivers License/State ID Number	800
Date of Birth	1,250
Credit Card Number	250
CVV of Card	50
Financial Account Number	2,750
PIN	100
Passport Number	50
Username	250
Online Account Password	200
Biometric Information	20
Military ID Number	10
Insurance Number	400
Medical Health Information	1,750
Medical Record Number	250
Security Question	100
PII Other	300

Privileged and Confidential



HAYSTACK REVIEW STATUS REPORT

PII Categories

5,000
4,500
4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
0
5,000
4,200
2,500
500
300
800
1,250
250
50
2,750
100
50
250
200
20
10
400
1,750
250
100
300
NAME
ADDRESS
SSN
TIN
FEDERAL ID NUMBER
DRIVERS LICENSE/STATE ID NUMBER
DATE OF BIRTH
CREDIT CARD NUMBER
CVV OF CARD
FINANCIAL ACCOUNT NUMBER
PIN
PASSPORT NUMBER
USERNAME
ONLINE ACCOUNT PASSWORD
BIOMETRIC INFORMATION
MILITARY ID NUMBER
INSURANCE NUMBER
MEDICAL HEALTH INFORMATION
MEDICAL RECORD NUMBER
SECURITY QUESTION
PII OTHER

QC Review	Overall
Batched QC Documents	2,250
Batched QC Docs Reviewed	1,800
Additional QC Docs Reviewed	750
% Complete	80.00%
% Accuracy Pre-QC	89.00%
Coding Changes	198
Batched QC Docs Remaining	**450**

QC Progress

80%
20%
Batched QC Docs Reviewed
Batched QC Docs Remaining

Reviewer Metrics	Reviewed	Rate/Hr.	QC Changes	Reviewer Metrics	Reviewed	Rate/Hr.	QC Changes
First Level Reviewer 1	600	32.50	13	First Level Reviewer 11	450	31.50	5
First Level Reviewer 2	750	27.00	13	First Level Reviewer 12	1,100	27.00	26
First Level Reviewer 3	700	23.00	10	First Level Reviewer 13	1,050	32.00	16
First Level Reviewer 4	550	24.00	6	First Level Reviewer 14	700	22.00	6
First Level Reviewer 5	1,200	24.50	22	First Level Reviewer 15	500	21.00	4
First Level Reviewer 6	1,000	22.50	19	First Level Reviewer 16	600	23.00	8
First Level Reviewer 7	800	21.50	22	First Level Reviewer 17	300	19.00	5
First Level Reviewer 8	550	23.00	6	QC Reviewer 1	1,300	27.00	1
First Level Reviewer 9	750	25.50	9	QC Reviewer 2	700	22.00	3
First Level Reviewer 10	400	23.50	4	QC Reviewer 3	1,000	25.00	0
				Total/Average	**15,000**	**24.85**	**198**

Normalization & Entity Deduplication

Classical Techniques

- SoundEx
- Nicknames, Common Abbreviations
- Human Review

Modern Techniques

- Template Matching
- Machine Learning Models

Confidence/Accuracy Scoring Focuses Human QC and Review of Final Disclosure List

Questions?