CONGRUITY | 360

CLASSIFY | 360 USE CASE | HEALTHCARE

Contents

CONGRUITY | 360

Introduction

Congruity360 is the current provider of data archiving services for a licensed health insurance subsidiary of a large national provider network within the United States (referred to as "Customer"). Customer suffered a substantial data breach that leaked both current and previous patients' personal and medical information. Beyond financial implications, Customer suffered significant reputational damage that compromised the trust it had built with its patient network.

Congruity360 was primarily tasked with storing sensitive data in a highly secure Air Gap environment 100% disconnected from the Customer network. Customer sought assistance to further their information governance and data security capabilities. This use case details how implementing the Classify360 solution allowed customer to achieve these requirements.

Business Challenge

At the inception of the project, one governing assumption was Customer's production data footprint was approximately 4 petabytes (PB) in 13 unstructured and 78 structured environments. The varied data locations included three legacy Isilon storage systems, SharePoint housed on an on-prem server, SQL database, and data in Microsoft Office365 and Microsoft Exchange. These 4 PB did not include lower environments used for User Acceptance Testing (UAT) and development. Data stakeholders existed across the legal, data security, and data governance departments. These disparate teams struggled to manage the organization's substantial data as a collaborative unit.

Customer sought to classify all data, perform disposition of non-compliant data, and properly store and archive all data per regulatory and business requirements. Future data storage could include cloud (e.g., Microsoft Azure), private hosting, and Air Gap (archival off network). Initially, Customer chose to focus on classifying and migrating data from the legacy Isilon systems.

Solution

Congruity360 proposed deploying Classify360, a data governance and rapid classification platform, to perform the classification of all data along with other information governance requirements. Congruity360 recommended a phased approach for implementing a full information governance program.

Congruity360 leveraged numerous discussions with key stakeholders to identify information governance requirements and priorities that lead to the following list of milestones and roadmap to complete the information governance and management deployment. Customer identified four "buckets" of secure data types on which to primarily focus classification efforts. Priority was granted to categories that fell under

significant, high-trust compliance requirements Customer was required to follow post-breach.

Phase 1 – Unstructured Data Auto-Classification

The highest priority for Customer was classifying "data at rest" for a 1 PB network file share based on the following four (4) top level categories:

1. Protected
2. Proprietary
3. Internal
4. Public

Applying these categories to data at rest helped manage data in flight and reduced the risk of end user behavior communicating sensitive data inappropriately. After considering DLP, endpoint protection, and other data security products, Customer identified these solutions and forced end users to manually classify data as they save it to disk. While this approach handled current data in use, Customer identified a large burden and overhead on the end user community. Additionally, these solutions did not address data generated before implementing these data security offerings, which goes back to the founding of Customer over a decade prior.

Based on these findings, Customer decided on an auto-classification process which leveraged a solution to perform the data classification in the background. Congruity360's Classify360 platform provides the ability to auto-classify all of Customer's data in the background. Classify360 has a library of robust connectors that can scan data at a content level wherever it resides in the corporate network, assuming the appropriate connectivity and permission are granted. Classify360 can leverage its supervised machine learning components to train models on the classification definitions provided by Customer. As Classify360 crawls new and legacy data, reports are generated based on classification assignment. On approval from Customer stakeholders, the Classify360 bi-directional connector framework then injects those classification definitions into all documents that support that capability. These classification definitions can be mapped to the existing DLP solution, reducing company exposure by capturing sensitive data in flight before it leaves the network.

Auto-Classification of Remaining Data Repositories

After the auto-classification pilot was signed off by Customer, Congruity360 continued with auto-classification of the remaining unstructured data repositories. Congruity360 and Customer discussed and identified the priority list for these remaining data repositories, and Congruity360 continued scanning the data and monitoring the auto-classification.

CONGRUITY | 360

Phase 2 – Full Information Governance Implementation

Granular Record Retention Classification Policy

This phase focused on implementing a more granular record retention classification policy. At the heart of Classify360 is the Information Engine, which leverages the in-place connectors to pull additional information from the content to achieve the more granular record retention classification policies. Customer needed major data loss prevention assurances; gaining an understanding of where risk data was and ensuring it was not exposed to breach was of utmost importance.

ROT Analytics and Actions

By implementing the full set of Classify360 capabilities, the following functions were available to solve Customer's information governance challenges:

- **Unsupervised Machine Learning –** Classify360 looked for patterns within Customer's data silos and recommend classification profiles with minimal to no input from the business.
- **Supervised Machine Learning –** Included as part of the auto-classification project, additional information pulled into the Information Engine was trained on for more granular and focused classification, based on the corporate record retention policies.
- **Risk Analytics –** Risk modeling identified content that can pose risk to the business, which is generally related to sensitive data such as PII, PCI, PHI, HIPPA, password protected, and corrupted files.
- **ROT Analytics –** Redundant, Obsolete and Trivial (ROT) data was identified and dealt with properly.
    - Redundant or Duplicates – Identified redundant or duplicate files within an individual data source as well as across all Customer's data silos. Customer leveraged Classify360's reporting to make informed and pragmatic decisions on data cleanup including prioritizing master versions and "system of record" repositories, as well as inserting reference links to single-instanced versions.
    - Obsolete – Identified content that adds no business value, such as data with no claim-related information that had not been accessed in the past 5 years, user files for departed employees, or other data past regulatory retention requirements.
    - Trivial – Identified trivial content based on predefined business rules by using the Classify360 best practices logic. Classify360 can identify non-compliant data (e.g., .mp3 files or iTunes libraries), downloads of the corporate MSDN subscription, and emails about the company holiday party.
- **Enhanced Reporting –** Provided a powerful and flexible reporting engine to create specific business needs templates as well as leverage Classify360's standard base reporting. The standard reports include predicting data growth, access control exposure needing lockdown, and owners of the growing data footprint.

Additional Classify360 Functionality Implementation

Classify360 provided a deep range of features to help map the enterprise data footprint to a company record retention policy. Built-in reporting provided insight to Customer. Feedback was solicited from data stakeholders and tweaks were made to perfect the modeling framework. In addition to creating customizable reports and generating email alerts, the bi-directional connector framework can perform many other actions.

After meeting with the key stakeholders, Congruity360 recommended the following actions to be most relevant to customer's information governance program:

- **Custom field metadata injection** – applying additional classification(s) into custom field metadata.
- **Smart link stubbing** – allows for the deletion of duplicate data, leaving behind a link to master document as an easy way to navigate to an alternative copy, reducing end-user impact while cleaning up duplicate content.
- **Deletion workflow** – allows the business to approve data deletion, which is recorded and leaves a defensible audit trail.
- **Access control lockdown** – removal of end user permissions to sensitive data.
- **Promote to Airgap** – an automated workflow to push data that cannot be deleted into the ultra-secure Air Gap archive.
- **Air Gap automation** – automates the manual Air Gap process to reduce Customer resource(s) required to archive requisite data to the Air Gap.
- **Redaction** – the ability to redact sensitive information from documents and structured data columns.
- **Cloud readiness** – based upon Customer's business rules, identify data eligible to be migrated to the cloud versus remaining on-premises or migrated to another secure data location.
- **Identity identification** – for handling the numerous Privacy Acts, identify all data for all data subjects proactively, ensure the ability to rapidly respond to data subject requests within the required timelines.

Phase 3 – R&D Information Source

The Classify360 connectors populated the Information Engine, which is a normalized and centralized location for all key information about the business. In addition to the main drivers for implementing a solid governance program, the gathered data was leveraged to enhance the business in numerous ways. By exposing Classify360's APIs, the business intelligence team was able to query the data to search for patterns and trends that could be leveraged and drive strategic decisions to better serve Customer's members.

Return on Investment

Implementing a solution to manage Customer's corporate data had many advantages. Turning that data into actionable information allowed the business to make key decisions to improve processes and mitigate risk.

The following benefits were recognized by implementing the Congruity360 solution suite:

Compliance

Risk reduction, fewer or smaller fines. By reducing the overall data footprint, preserving only data required by regulatory or business demands, applying granular retention policies to the various types and quantity of data, and having a well-defined and documented information governance process, Customer was able to meet various regulatory and business compliance requirements. Data could be identified quickly and confidently, which reduced resources, operational costs, and potential fines in the event of a regulatory issue. Having data identified and classified allowed Customer to reduce the reactive time-consuming approach to a more proactive approach, saving time and money.

- Privacy Act – found all data for data subject requests and handled within each Privacy Act's requirements. With various Privacy Acts being introduced into or already passed by several state governments, Customer was able to identify all data for each data subject. Until the U.S. government passes a federal-level privacy act, most enterprises must handle many different acts put forth at the state level. Leveraging Classify360's flexible modeling, identity identification, and classification engine enabled customer to meet these requirements in a timely manner.
- Security – less data exposed after classification and proper storage. After scanning and classifying all data, customer disposed of any data not required to keep per the numerous regulatory, industry, and business requirements – resulting in less data exposure risk. Any sensitive data that was still on unsecured data storage was identified and moved to proper storage, whether on-premises, cloud (e.g., Microsoft Azure, AWS, etc.), private hosting, or Air Gap. Lastly, Classify360 combined with DLP, end detection, and other data security solutions increased customer's data security and governance capability.
- Technology – less technology (primarily storage), less cost related to technology and employees (support, maintenance, help desk). In most environments, enterprises experience a 15-25+% data reduction, thus leading to the elimination of old and costly storage, potentially less compute resources (e.g., CPUs, RAM), and focusing scarce IT resources to other technology initiatives.
- Records management – increased and improved capability maturity model (CMM) associated with creating and managing records. Once all data is scanned, indexed, and classified, Customer methodically and efficiently categorized the data into a proper and useful record management protocol.
- Operational improvement – as more data is classified, less time is needed to find specific data. Without enterprise data classification, much time is wasted performing data searches for employee or business reasons across the various company applications or data storage areas. Full data classification allowed employees and systems to connect with all the right data more quickly, reducing wasted time and increasing overall operational effectiveness.
- Change management – improved behavior related to data savings (i.e., unstructured) with higher percentage of data stored in correct location. Until an effective information governance program was place and addressed the ad hoc storing and saving of data, Customer had no ability to introduce, change, and manage data behavior. After all data was scanned and classified, the Customer can now properly manage data behavior, first by identifying what type of data is stored where, and second by generating sustainable employee information governance

procedures. If employees keep repeating poor information governance practices by storing data in incorrect locations, then Customer can generate apply change management and organizational behavior training to reduce and hopefully eliminate the subpar governance practices.

- R&D – assisted with improving R&D processes and predictive health patterns. The Customer enterprise had lots of data that could be farmed to produce more effective health results. The Classify360 classification engine and enterprise search was leveraged to find any data, documents, or other files based on a wide variety of classes and terms. Customer leveraged this knowledge by improving current and creating new health solutions and recommendations, both improving members' lives and making a better healthcare network.
- Enterprise growth –savings from soft and hard cost reduction was shifted from operating the business to growing the business. By reducing operational costs, Customer reinvested the savings to expand offerings to members and be more financially competitive in the healthcare marketplace.
- Thought leadership – Customer demonstrated thought leadership among its network peers. Deploying an industry leading information governance solution leading to effective and valuable changes in the healthcare market gave Customer the opportunity be the exemplar within the local community as well as the wider healthcare industry.

The following table contains sample metrics gathered from both implementations of Classify360 and Gartner's analysists assigned to Congruity360.

	Description	**Potential Savings**
Data Storage	**An average deployment of Classify360 will identify 22.5% of corporate data that can be defensibly deleted through a process of identifying duplicates and content that add no business value. Focusing on the larger data silos first these returns should come into effect by the end of the first year. The savings is based on a 4 PB footprint and an internal enterprise storage cost of $300 per TB per month.**	**$3.31M returned to the storage budget annually.**
Privacy Act Compliance	**Current "collaboration" approach to satisfy data subject access requests (DSAR) for 5,000 employee company costs about $10,000 FTE hours to comply. An average of 100 data subjects per year would cost $1M.**	**Automating the process would save minimally 50% if not higher. This would translate to $500K savings per year.**
Data Breach	**The average global cost of a data breach is $3.86M, and while the average U.S. cost is $7.91M.**	**$148 per Customer member or past member exposed.**
Brand or Reputation	**For a company with $1B annual revenue, brand or reputation can be damaging.**	**A small 1% revenue loss equates to $10M.**
Fines	**The fines can originate from local, state, or federal government entities, courts, or other. If an enterprise had annual fines of $10M total due to bad information governance, a minimal 10% savings.**	**$1M per year.**

CONGRUITY | 360

Conclusion

Customer's large-scale, ongoing deployment of Classify360 has already resulted in a myriad of benefits:

- 1.46 billion Isilon documents and 80 million Exchange documents (in a matter of weeks, not months) and 250,000+ Microsoft O365 and SharePoint documents (this portion of classification is in its infancy) have been processed to date
- Cross-departmental collaboration and data management is now possible, allowing true data governance to occur across the organization
- Hyper-specific workflows ensure compliance is maintained with both legacy and newly created data with minimal manual effort
- ROT data has been moved out to cheaper storage in the cloud, saving money and operational overhead; data older than 5 years with no business value yet protected by regulations is automatically pushed to archive storage
- Customer not only took steps to fulfill its court-ordered financial obligation by investing in Classify360, but also implemented a platform providing ongoing governance and risk mitigation to reduce the impact of future breaches and provide ancillary cost savings organization-wide

At a time when it's not a question of "if" but, rather, "when" an enterprise's next data breach will occur, Customer is at peace knowing its highly regulated and risky data is protected with the utmost security and caution. Healthy, ongoing retention policies have been put into place with automated workflows, monitored by a team of cross-functional data stakeholders. Customer is now using its breach experience to educate its peers within its healthcare network, sharing the tactics and solutions that have enabled them to come out of a critically difficult situation fully prepared for future success.