CONCENTRIC

DATA RISK REPORT

Q2 2020

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This is the Q2 2020 edition of Concentric AI's quarterly Data Risk Report. This report is based on live data captured by the Semantic Intelligence™ solution to reveal how organizations create, use, and manage data.

The last few months were not, by anyone's definition, business as usual. Public health concerns led to economic dislocations not seen for a generation. As organizations scrambled to keep workers connected and communicating, they also confronted significant changes in the threat landscape. Cyber criminals preyed on public health fears. Employees unaccustomed to remote work found themselves collaborating on-line and sharing more data. IT professionals re-prioritized initiatives to focus on work-from-home programs and technologies. We've re-prioritized as well: in this report we'll focus on possible security impacts of work-from-home practices with a focus on unstructured data.

The number of overshared files grew by 52%.

EXECUTIVE SUMMARY

Over 80% of an organization's data is unstructured[1], meaning it's embedded in the millions of financial reports, corporate strategies documents, source code files, and contracts created by CFOs, general managers, engineers, and lawyers every year. But to an IT security professional, unstructured data is still a shapeless lump of clay – unseen, unexplored and insecure. Using advanced AI capabilities, Concentric processed 50 million unstructured data files from companies in the technology, financial, and healthcare sectors to create this report. By categorizing documents, evaluating business criticality, and accurately assessing risk, we give visibility and shape to unstructured data, helping security professionals make more informed decisions and develop policies that reflect the realities of expanded remote work and collaboration.

Unstructured Data Trendlines

TOTAL FILES PER ORGANIZATION	10.2 M	Up from 9.9 M
BUSINESS-CRITICAL FILES PER ORGANIZATION	2.8M	Steady
OVERSHARED FILES PER ORG	289,000	Up 52%
OVERSHARED FILES PER EMPLOYEE	105	Up from 38

[1] *https://en.wikipedia.org/wiki/Unstructured_data*

NEW THIS QUARTER

Our analysis focused on the potential security impact of expanded work-from-home practices. We see evidence of elevated data risk. Since our last report:

- The number of overshared files detected grew by 52%.
- Organizations have an average of 289,000 overshared files. That's 105 files per employee (up from 38 files per employee in Q1 2020).
- Link-based sharing is up to 45,000 documents per enterprise (from 35,000 documents per enterprise in Q1 2020).
- Profiled enterprises laid off 5% of their workforce or about 150 employees each. This raises the specter of insider threats although we currently cannot quantify the risk.

While proving definitive causal relationships is difficult, our observations suggest recent changes to how we work and interact have increased data risk.

CONSISTENT WITH Q1 2020

- Unstructured file counts remained steady at ~10 million/organization
- 90% of business-critical documents are shared outside the C-suite
- Over 11% of all business-critical files can be seen by internal or external users who should not have access

NEW ANALYSIS CAPABILITIES

Concentric now supports 91 business-critical categories, up from 51 in Q1 2020

UNSTRUCTURED DATA SECURITY MORE CRITICAL THAN EVER

End users have an outsized impact on unstructured data security. Sharing practices, access control decisions, and classification accuracy determine whether a file is secure or at risk - and these file-by-file decisions are beyond the reach of centralized security infrastructure. As our data shows, work-from-home makes managing unstructured data risk even more critical.

But protecting this data is unfamiliar territory for many security pros. Millions of files need to be scanned to find those that are both business critical *and* inappropriately shared. It's no small feat, and it's the crux of the unstructured data security problem: out of the 10.2 million files an average enterprise has, how can we know which files are overshared without overwhelming IT teams with false positives?

TAKING SHAPE

Unstructured data is diverse, both in form and content. Many files are mundane and represent no real threat if overshared or stolen. Others contain information critical to the business. So, the first – and perhaps most difficult – task is to determine which documents we should worry about.

Concentric used sophisticated deep learning techniques to categorize over 50 million files from companies in the technology, finance, and healthcare sectors. We discovered that typical organizations have over 90 different types of business-critical information hidden in unstructured data (grouped here for clarity):

PRODUCT

Bills of materials, source code, design documents, and test plans

At stake: intellectual property loss, product liability, customer anxiety, strategic disclosure

FINANCIAL

Bookings, income, forecasts, pricing, invoices, trading, and tax filings

At stake: insider trading violations, compliance, loss of competitive advantage

LEGAL

Non-disclosure agreements, contracts, purchase agreements

At stake: civil lawsuits, loss of favorable supplier terms, other legal liabilities

HUMAN RESOURCES

Offer letters, stock agreements, consulting contracts

At stake: employee satisfaction, private information, higher costs

SALES

Requests for proposals, quotes, customer strategies

At stake: lost business, strategic disclosure, sales team dissatisfaction

PARTNER

Mergers and acquisition docs, partner agreements

At stake: damage partner relationships, sink acquisition initiatives, encourage insider trading

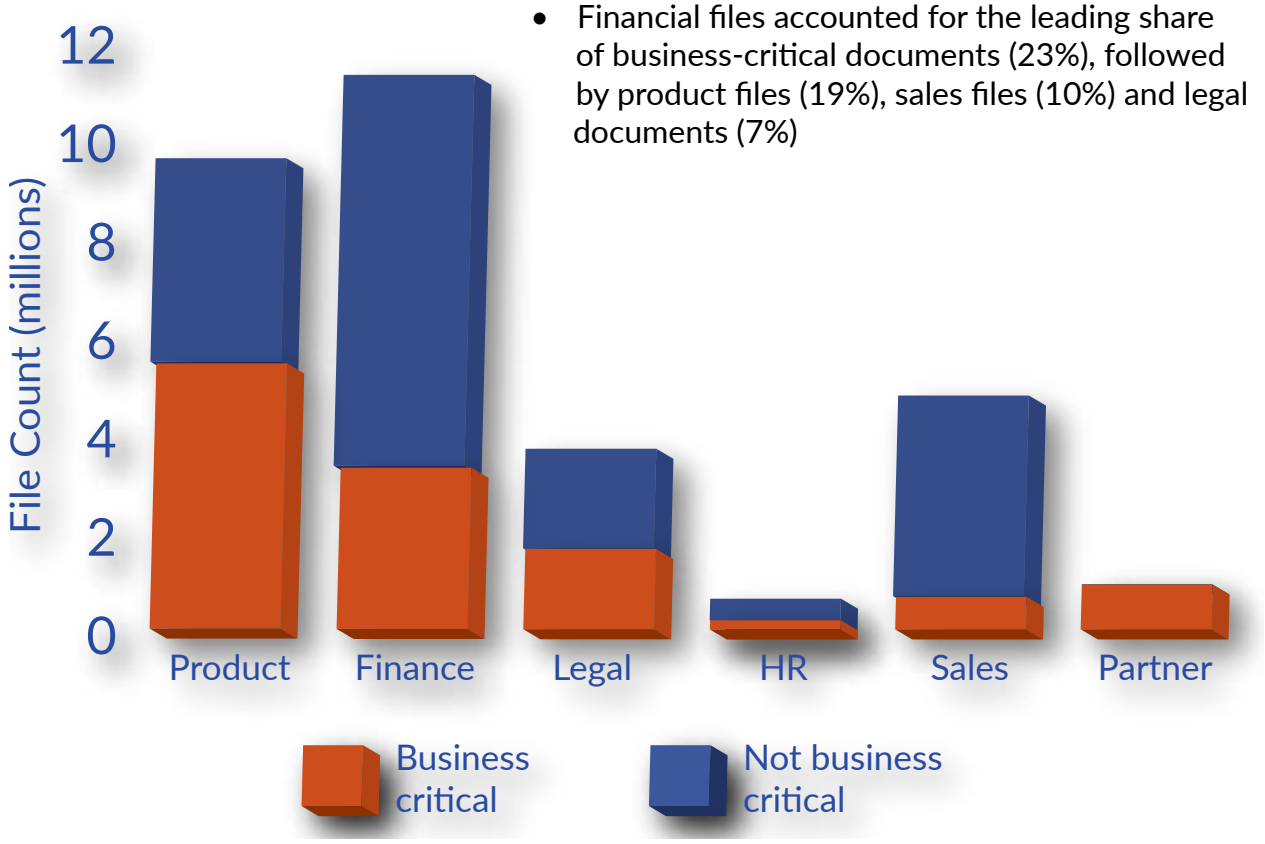
Concentric autonomously assigns business-critical files to one of over 90 categories.

SEEKING MEANING

Concentric evaluates business criticality based on contextualized content, file ownership, document meta-data, presence of personally identifiable information, and peer file comparisons. Business criticality is, of course, vital to security assessment. These are the files that must not be overshared.

Here's what we learned:

- Nearly 27% of an organization's unstructured data is business critical (2.75M million files on average per organization)
- On average, each employee is responsible for 449 business critical documents
- Financial files accounted for the leading share of business-critical documents (23%), followed by product files (19%), sales files (10%) and legal documents (7%)



FINDING ANSWERS

> Sharing a contract with the legal team might be appropriate. Sharing it with the engineering team might not. It depends.

Assessing risk – even with a fully categorized set of files - is a deceptively complex task. Appropriate sharing depends on the meaning and function of the document itself.

Peer file comparisons provide accurate risk assessments by leveraging the wisdom of users creating and managing similar files. We compare each document's security characteristics to those of its peers to reliably identify oversharing – inside or outside the organization. We evaluate:

INTERNAL SHARING

Is internal user sharing consistent? This is tough to spot without peer file analysis – and it's critical for security.

MISCLASSIFIED FILES WITH PII

Is this document marked to indicate it contains PII? Classifications for PII can also help a DLP solution fence in PII to maintain compliance.

MISCLASSIFIED CONFIDENTIAL FILES

Has this document been properly classified? Document metadata, such as a "confidential" tag, is routinely used by other security solutions to enforce policy (e.g. a DLP solution uses a tag's setting to block a document).

SHARING WITH GROUPS

Do peer files allow similar group access?

SHARING WITH EXTERNAL USERS

Are similar documents shared with external users? Are they the same external users?

Risk Scenarios in Business-Critical Files

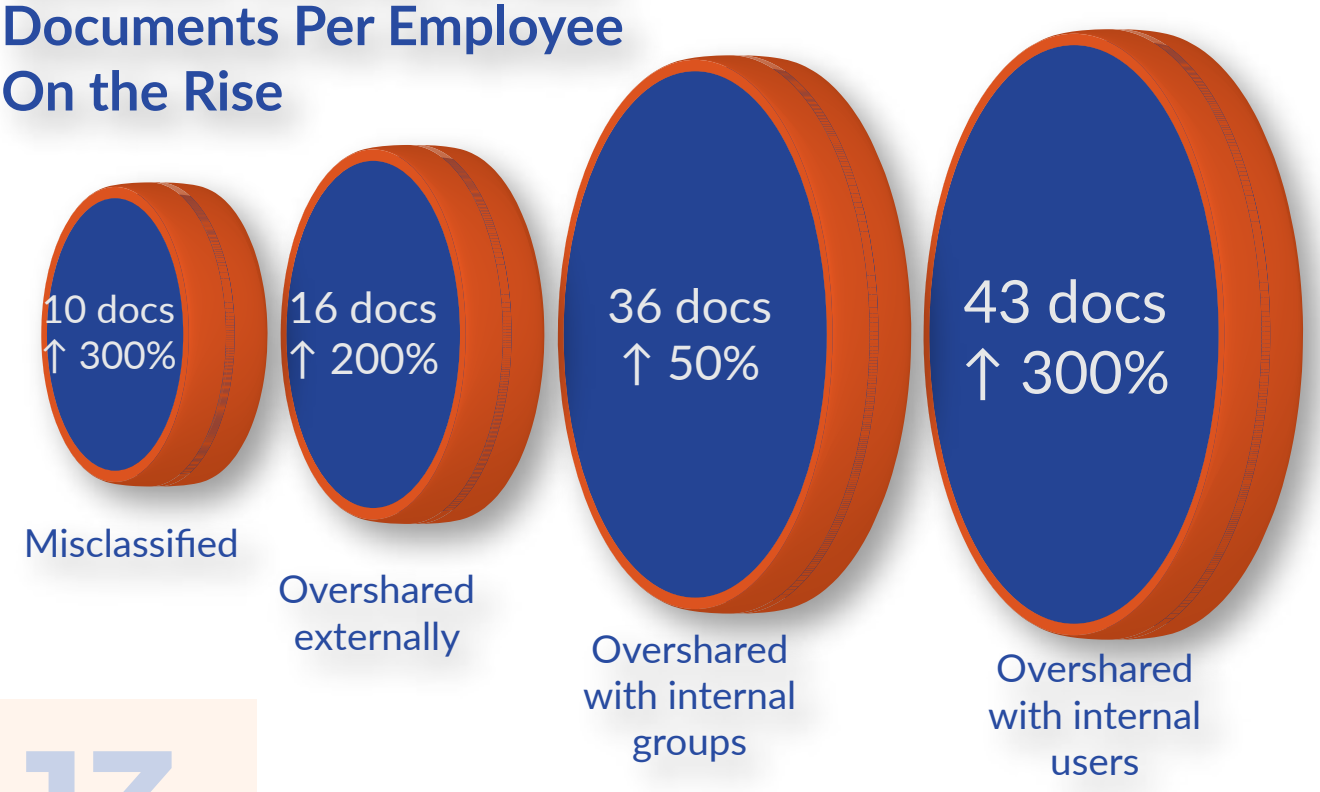
EXTERNAL USER ENTITLEMENT MISMATCH	Documents inappropriately shared with external users
GROUP ENTITLEMENT MISMATCH	Sensitive data shared erroneously with groups
INTERNAL USER ENTITLEMENT MISMATCH	Sensitive data shared erroneously with internal users
MISCLASSIFIED DOCUMENT	Confidential documents that are not properly marked/classified
UNCLASSIFIED CONTAINING PII	PII data in documents that are not properly marked/classified

FINDING ANSWERS

We discovered some surprising results:

- 11% of an organization's business critical data is overshared
- On average, each organization had 289,000 files at-risk due to oversharing (105 files per employee) up from 190,000 in Q1 2020 (38 files per employee)
- Documents in the product and finance categories accounted for 42% of the total number of overshared documents
- 84% of the at-risk files were overshared with users or groups within the company (up from 75% in Q1 2020)
- 139,000 business-critical files were erroneously classified and accessible by employees who should not have access to them

Business-Critical, At-Risk Documents Per Employee On the Rise

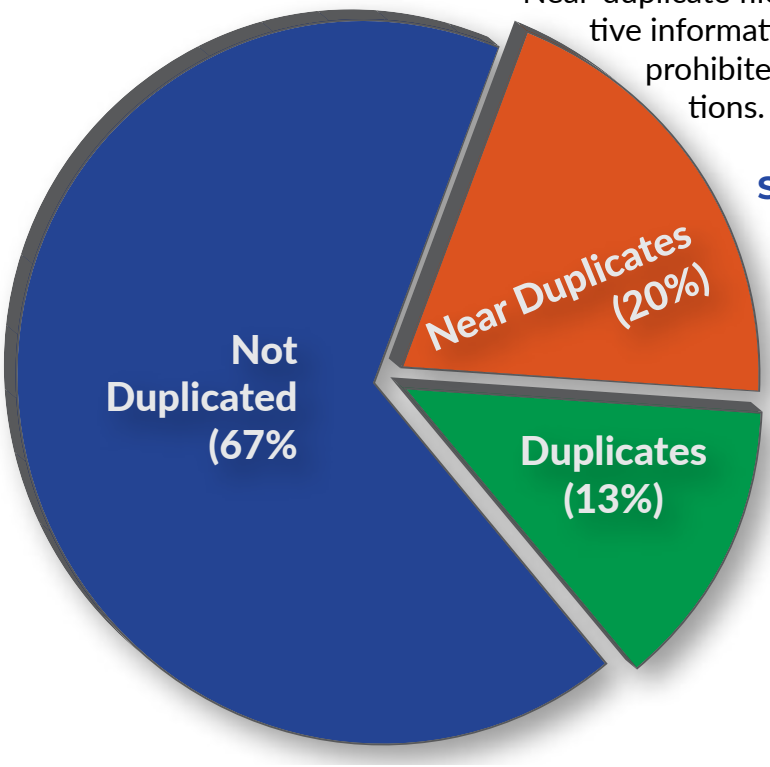


DANGEROUS PATTERNS

After reviewing the data, we noted some patterns that seemed to reoccur across companies, regardless of sector or company size:

NEAR-DUPLICATE FILES

Nearly 1 in 3 files we processed were identical or nearly identical. Near-duplicate files create multiple variant copies of sensitive information, often with insecure file permissions, prohibited locations, or improper file classifications.



SHARED WITH EVERYONE

A shocking number of business-critical files were shared with everyone in the company. We found 62,842 such files, containing everything from source code (4,500 files) to contracts (5,028 files) to offer letters (1,072 files).

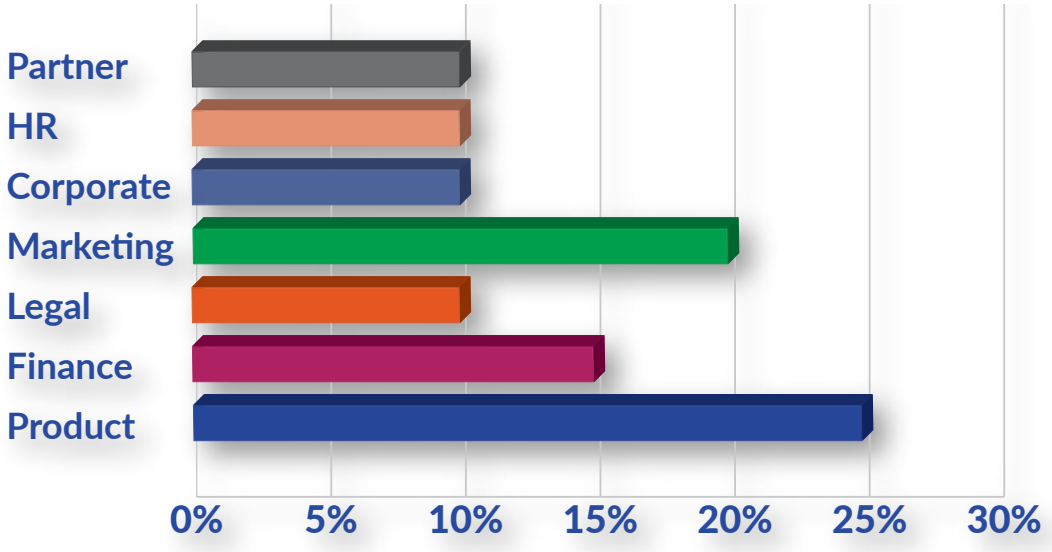
DANGEROUS PATTERNS

INTERNAL OR EXTERNAL OVERSHARING

Of the 289,000 at-risk files, 84% were overshared with users or groups outside the proper team or department. These issues are nearly impossible to identify without peer document comparisons. We found 19,567 source code files, 15,745 design documents, 15,355 purchase agreements, 4,957 tax filings, 21,675 stock agreements, and 6,903 offer letters among these files.

PERSONALLY IDENTIFIABLE INFORMATION (PII)

PII is of interest due to privacy concerns and regulatory requirements. Increasingly, security teams use document metadata to flag PII and control document sharing and transfer. Nearly 21% of all documents containing unstructured data contained PII and were not marked appropriately.



Documents Containing PII by Function

OVERSHARING IS A MODERN ENTERPRISE REALITY

We've provided no shortage of statistics in this report. Statistics, sometimes, don't convey what's really happening on the ground. To help keep it real we offer a few specific incidents that show just how easily oversharing happens.

"COMMON" FOLDER

A mid-sized financial services company uses a common folder to share non-critical documents with all employees. Out of the 100,000 files in that folder, 1,000 contained proprietary trading information or contracts. Similar documents located elsewhere in the company were highly restricted.

ENGINEERING COLLABORATION

At one high-tech firm, source code and design documents were routinely shared with everyone in the company. We found over 12,000 overshared files.

OFFICE 365

In another high-tech firm, a user slightly modified a corporate strategy document, moved it to an on line Office 365 location, and shared it with two external users. By identifying a near-duplicate file we were able to identify oversharing.

INTERDEPARTMENTAL OVERSHARING

At another financial services company, our research uncovered a sales contract shared with the Engineering and HR groups. Mis-configured access settings are common but can be as hard to find as the proverbial needle in the haystack.

CONCLUSION

The cloud, mobile computing, network virtualization, and a host of other technologies have remapped the IT landscape. The pandemic - by changing how we work and collaborate - is remapping IT once again.

It would be foolish to forecast how the world will look next year (or even next quarter). But one thing is clear: the need to secure our unstructured data has never been greater.

CONCENTRIC

Eighty percent of corporate data is found in the files and documents employees create and use every day. Concentric discovers and categorizes this unstructured data to protect intellectual property, financial documents, PII/PCI content and proprietary business information (strategy plans, product roadmaps, contracts, blueprints) wherever it's stored. Our Semantic Intelligence™ solution uses deep learning to develop a semantic level understanding of a document's content to discover business sensitive data, surface risks, and remediate issues without relying on upfront rules or complex configuration.

www.concentric.ai
Twitter: @IncConcentric
LinkedIn: linkedin.com/company/concentricinc

4340 Stevens Creek Blvd
Suite 112
San Jose, CA 95129

Vatika Business Centre,
Cessna Business Park
5th Floor, Embassy Signet
Kadubeesanahalli
Outer Ring Road
Bengaluru, India 560103