DATA RISK REPORT

2H 2023

AUTONOMOUS DATA SECURITY

CONCENTRIC

AUTONOMOUS DATA SECURITY

TABLE OF CONTENTS

© Concentric 2024 | v1.0 | All rights reserved

> Concentric autonomously assigns data to one of over 276 categories. Over 175 of those categories are business-critical.

EXECUTIVE SUMMARY

Thus is the 2H 2023 edition of the semi-annual Data Risk Report published by Concentric AI.

To an IT security professional, data is still a shapeless lump of clay – unformed, unseen, and insecure. Most enterprises lack visibility into where their sensitive data is, much less where the risk is to the information from entitlements, sharing, permissions, activity etc.

Using advanced AI capabilities, Concentric processed 1 billion structured and unstructured data records / files from companies in the technology, financial, energy and healthcare sectors. This report gives shape to the state of risk to data in the real-world by categorizing the data, evaluating business criticality, and accurately assessing risk. Accuracy is critical – it's the difference between effective protection and alert fatigue caused by thousands of false positives.

All results in this report were based on live data analyzed by the Concentric Semantic Intelligence™ solution autonomously derived to reach our conclusions. Here are a few things we learned and how the data compares to 1H 2023

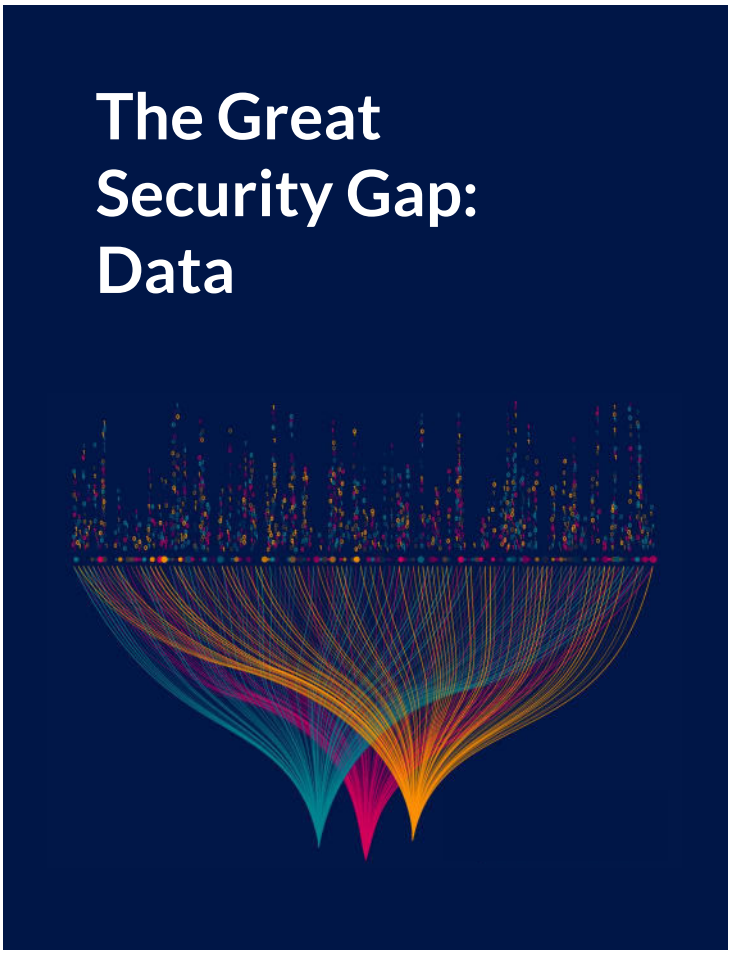
- Concentric identified 276 biz critical categories (up 10% from last cycle)
- Nearly 32% of an organization's unstructured data is business critical (meaning its distribution should be controlled)
- 90% of business-critical data are shared outside the C-suite
- Over 15% of all business-critical data files/records are at risk from oversharing, erroneous access permissions and inappropriate classification and so can be seen by internal or external users who should not have access

15TB	Data files/records per avg enterprise
5TB	Biz critical data per avg enterprise
0.75TB or 443 data records per employee	At risk biz critical data per employee (up from 402 in 1H 2023)

OVERALL

- Risk due to oversharing continues to trend up – a 12% HOH increase. Some of it is attributable to the increase in sample size but adjusting for that, there has been an increase in risk due to sharing. The data proves that inspite of all the cyber security investments, data remains a vulnerable threat surface
- On average, each organization had 1.1M data records at-risk due to oversharing (443 files/records per employee) up from 802K in 1H 2023 (402 data files/records per employee)
- Link based risky sharing is up to 100K documents per enterprise (from 81K documents per enterprise in 1H 2023)

	1H 2023	2H 2023	Delta
Total data analyzed	500TB	1PB	+200%
Biz Critical data per enterprise	5M	6M	+20%
Data at risk from Oversharing	802K	1.1M	+38%
Data at risk per employee	402	443	+10%
Avg Employees per customer	1995	2501	



Today, perimeter control and database protection products get the lion's share of security spend. Firewalls, access control frameworks and cloud access security brokers (along with many other security solutions) are large, established product categories. Enterprises have options.

But the options to protect data aren't nearly as focused or effective. It's easy to see why: scanning PBs of data to accurately identify those that are both business critical AND inappropriately shared is no small feat. This, then, is the crux of the data security problem: out of the 15.2 million data records an average enterprise has, how can we know which data are overshared without overwhelming IT teams with false positives?

Taking Shape

Data is diverse, both in form and content. Many files/data records 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 1 billion files/data records from companies in the technology, finance, energy, chemicals, university and healthcare sectors. We discovered that the average organization has over 276 different types of biz critical categories hidden in its unstructured data (grouped here for clarity):

Taking Shape

Product

(representative categories include bills of materials, source code, design documents, and test plans) - overshared product files can result in a loss of intellectual property, increased product liability, customer anxiety, and strategic disclosure.

Financial

(representative categories include bookings, income, revenue forecasts, pricing documents, invoices, trading, and tax filings) - oversharing these files can result in insider trading violations, compliance liabilities, and loss of competitive advantage.

Legal

(representative categories include non-disclosure agreements, contracts, and purchase agreements) - exposure of a sensitive legal file can expose the company to civil lawsuits, loss of favorable supplier terms, and other legal liabilities.

Human Resources

(representative categories include offer letters, stock agreements, and consulting contracts) - oversharing HR files can harm employee satisfaction, reveal private information, and driver higher costs

Sales

(representative categories include requests for proposals, quotes, and customer strategies) - exposing sales files can result in lost business, strategic disclosure, and sales team dissatisfaction.

Partner

(representative categories include mergers and acquisition documents and partner agreements) - losing partner files can damage partner relationships, sink acquisition initiatives, or encourage insider trading.

Source Code Design Docs Test Plans Roadmap

Pricing Tax Filings NDA Stock Agreement Consulting Agreement

Trading Income & Booking Revenue Forecasts Bookings Bill Of Materials

Offer Letters Contracts M&A RFP Purchase Agreement

Invoice Quotes Opportunities Partner Agreement

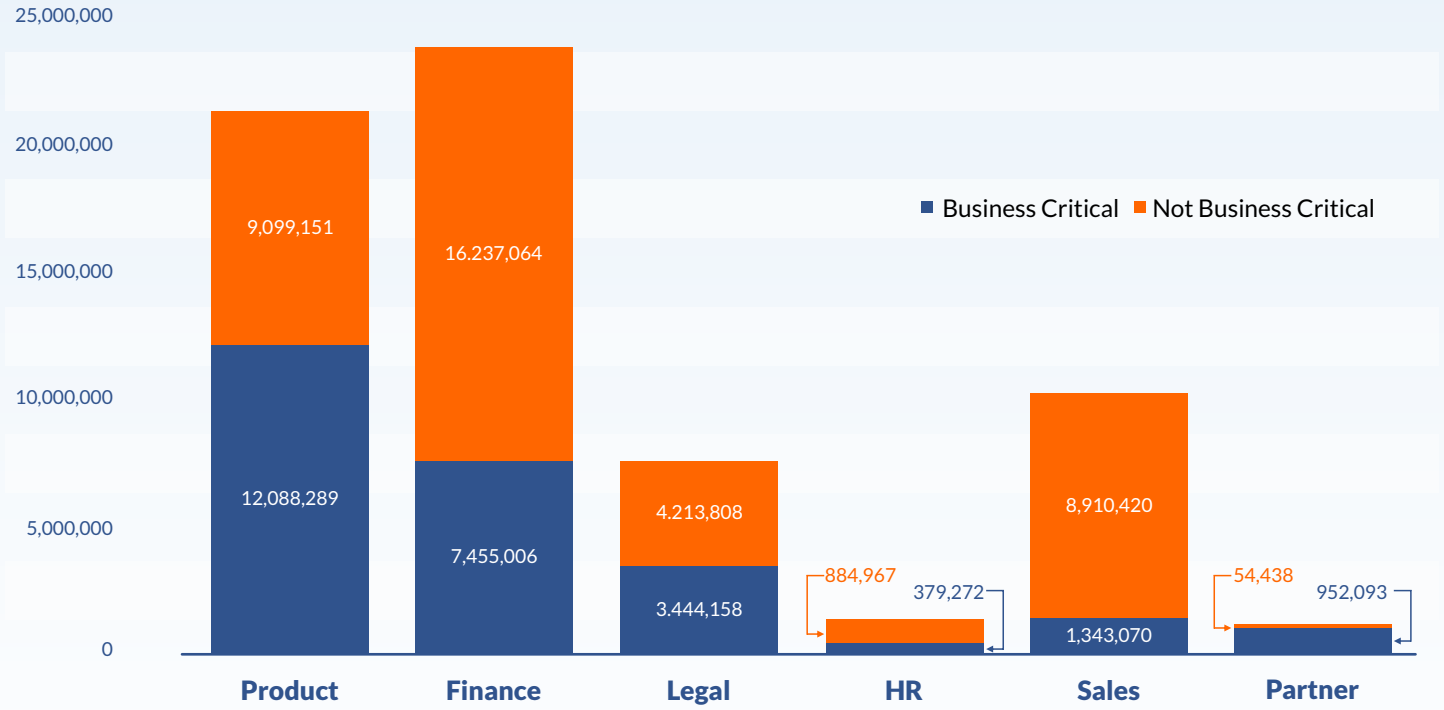
Seeking Meaning

Once categorized, Concentric evaluated each file for business criticality based on a variety of factors including , contextualized content, file ownership, document metadata, presence of personally identifiable information, and peer file comparisons. Business criticality is, of course, a vital piece of the puzzle. These are the files that must not be overshared.

HERE'S WHAT WE LEARNED:

- Nearly 32% of an organization's unstructured data is business critical (6M million files/ data records on average per organization)
- On average, each employee is responsible for 2506 business critical documents
- Financial data accounted for the leading share of business-critical documents (24%), followed by product files (22%) and sales files (10%) and legal documents (8%)

Data by Category



Finding Answers

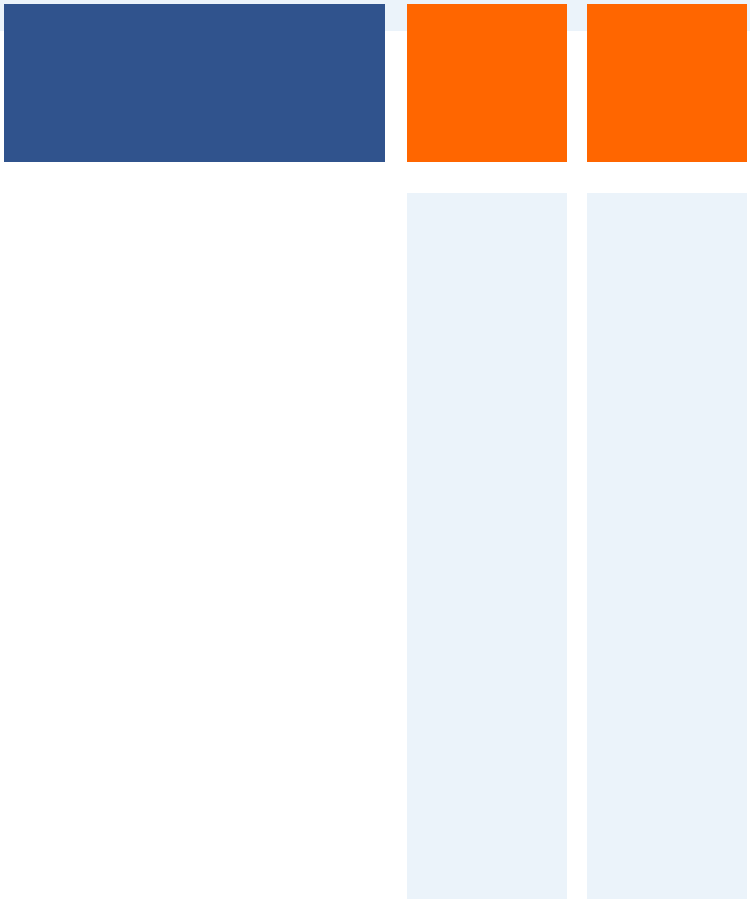
Assessing risk – even with a fully categorized set - is a deceptively complex task. Appropriate sharing depends on the meaning and function of the document itself. Sharing a contract with the legal team may be appropriate. Sharing it with the engineering team might not. It depends. Rigid rules-based risk assessments can be wildly inaccurate, especially when applied to policies governing intra-company file sharing.

To gain an accurate picture of risk, our analysis starts with the categories developed in the previous steps. We compare each document's security parameters to those of its peers. Using peer data configurations as a benchmark we can reliably identify oversharing – especially between employees at the same company.

We check for:

- **Sharing with external users** – are similar documents shared with external users? If so, are they the same external users?
- **Sharing with groups** – do peer files allow similar group access?
- **Sharing with internal users** – is internal user sharing consistent? This is tough to spot without peer file analysis – and it's critical for security.
- **Misclassified confidential data** – 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) .
- **Misclassified data containing PII** – is this document marked to indicate it contains PII? Classifications for PII can also help a DLP solution fence in PII to maintain privacy and compliance.Wrong location
- **Anonymous link sharing**
- **Sharing with personal email accounts:**

COMMON SCENARIOS THAT INCREASE RISK



DANGEROUS PATTERNS

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

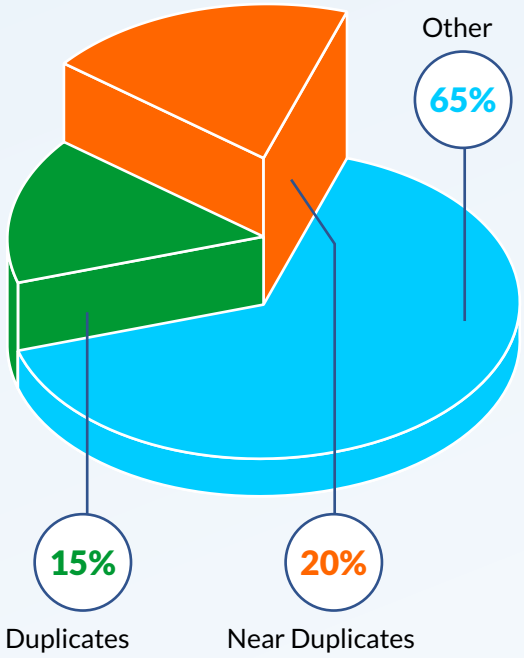
Near Duplicate files.
1 in 3 files we processed were identical or nearly identical. Near Duplicate files create multiple variant copies of sensitive information, often with different (and incorrect) file permissions, prohibited locations, or improper file classifications.

Shared with everyone
A shocking number of business-critical data were shared with everyone in the company. We found 160,078 such files/ data records

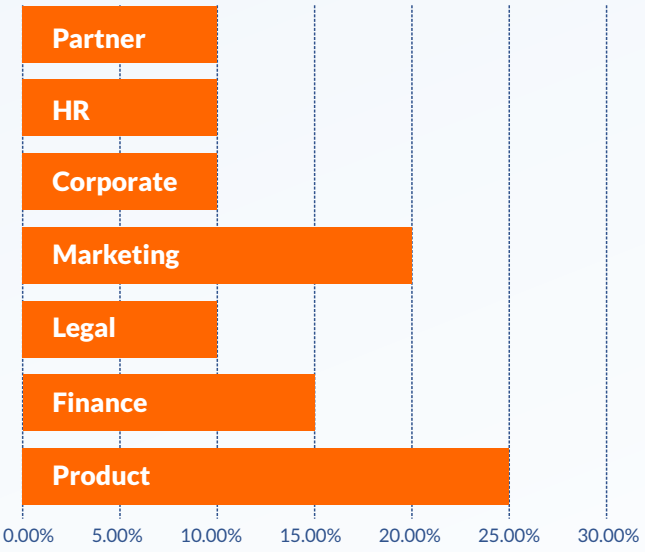
Internal or external oversharing
Of the 1.1M at-risk data, 83% were overshared with users or groups outside the proper team or department. These issues are nearly impossible to identify without peer data comparisons.

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 25% of all documents containing unstructured data contained PII and were not marked appropriately.

Duplicates and Near Duplicates



PII as percentage across categories



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.

Risky sharing outside the company
Bob in Finance at an energy firm shared sensitive ITAR protected data with a friend

Not enough access controls on IP or PII data, ie. too permissive
Judy in HR at a financial services company ended up having access to our highly-sensitive financial intellectual property

Risky sharing to personal email
The CFO at a high tech company needed access at home and she sent the confidential 2023 budget to her Hotmail account

Sensitive data in the wrong location
At a healthcare firm, we found healthcare documents wth PHI stored in Office365, Google Docs, and Dropbox when they wanted it securely stored on AWS S3

Inappropriate classification
At a financial services firm, we found mortgage document that wer not classified and consequently open to almost everyone in the IT department

Conclusion

Data sharing's transformational impact on businesses is indisputable. The productivity of pre-network paper-based communications pales compared to today's instantaneous electronic world. Hot new technology trends – like cloud computing and corporate digital transformations – all advance one fundamental goal: more, and more effective, data sharing.

But sharing has its dark side. The documents that used to fill physically secure filing cabinets can now be shared instantly, and with anyone. Businesses risk oversharing confidential sales strategies, need-to-know M&A plans, and sensitive personnel information with people who shouldn't have access. Autonomous and intelligent technologies are now able to find, categorize, and assess large bodies of structured and unstructured data for better data security posture management and security.