USE CASE

Semantic Intelligence™ for Data Deduplication and Data Retention

Corralling data in an increasingly connected world can feel like you're trying to bring order to the Wild West. With sensitive data scattered everywhere and growing exponentially, it's not uncommon for organizations to have multiple copies of the same file or to be hanging onto very old data that they no longer need.

In fact, a recent Concentric AI Data Risk Report analyzed unstructured data records from companies in the technology, financial, energy, and healthcare sectors and found that one in three files were duplicates or near duplicates.

Duplicate and near duplicate data is often improperly classified, has incorrect permissions, and is stored in unauthorized locations, which puts it at significant risk. How do you adhere to regulations such as the Privacy Act and protect data from unauthorized disclosure when you have no idea how many copies you have out there to begin with?

Retaining multiple copies increases storage costs and cyber insurance premiums, creates versioning chaos, and impacts productivity. You've got sales presenting outdated content to prospects, finance updating the wrong purchasing agreements, and legal struggling to stay on top of the most current SLAs.

Nothing Golden About These Oldies

Your company's issued guidelines on how long data should be retained, but no one's following them and your repositories are full of data records that are way past their sell-by date.

The problem is, you don't have visibility into how old the data is or when it was last modified. No one's willing to hit "Delete" because they don't want to risk accidentally getting rid of records that the company still needs.

So here you are, hanging onto data that should've been deleted or archived years ago, and the storage costs are giving you heartburn.

You need to know exactly what you have so you can confidently ditch your data dead weight.

Sticker Shock

Cyber insurance premiums are based on how much sensitive data an organization has, but since you haven't done discovery, you have no idea.

Because you can't provide your insurer with a concrete number, they're just assuming it's all sensitive—and charging accordingly.

It's Time for Your Data Glow-Up

Concentric AI's Semantic Intelligence platform is unique in its capability to discover and categorize data. While other tools can identify the same social security number in multiple data records, they can't tell you if these records are duplicates.

Our AI doesn't just look for patterns or phrases, it reviews entire data records, so it understands context. Not only can we identify when there are multiple instances of a social security number, but we can also tell you if it's duplicate or near duplicate data—or if it's showing up in multiple places for legitimate reasons, for example because it's required in both finance and healthcare data records.

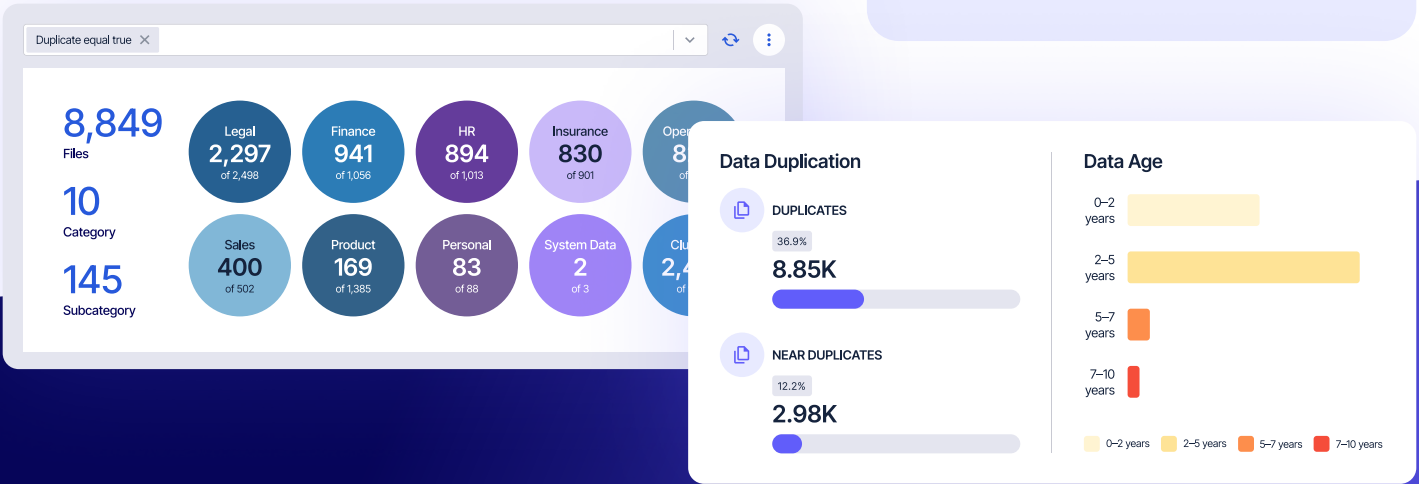
Once we've discovered and categorized your data, we'll help you clean it up. Within the platform, data deduplication is as easy as checking a few boxes in the Content Explorer dashboard. You'll use the Similarity filter to identify data by category and subcategory and then select from the Create Policy menu to remediate—you can delete, copy, or move data to a destination of your choosing.

We'll also help you set up and enforce automated data retention policies, so your data is stored only where it should be, moved if it isn't, and not kept longer than it's needed.

Less Clutter, More Clarity

Move over, Marie Kondo. With Semantic Intelligence, you get the peace of mind that comes from knowing your data's gone from crime scene to deep clean.

- **Cut costs:** Spend less on storage and watch those cyber insurance premiums drop.
- **Focus on productivity:** No more wasting time hunting for the right file.
- **Ensure compliance:** Easily set data retention policies in the platform.
- **Stop the spread:** Enable better data hygiene with a solution that discovers all your duplicate, near duplicate, and stale sensitive data.



About Concentric AI

Concentric AI is intelligent data security made easy. Our platform discovers, classifies, monitors risk, and protects data across cloud and on-premises environments. With AI-driven insights and co-managed services, we help businesses reduce risk, streamline compliance, and eliminate duplicate data records—transforming data from a liability into a well-managed asset. Contact us today to schedule a live demo and see for yourself how Semantic Intelligence helps you discover and deduplicate your most sensitive data.

Schedule a Demo