ACTIVENAV

Cleansing and Organizing Unstructured Data Stores in Preparation for Migration

How the Metropolitan Water District of Southern California came to understand and organize their unstructured data sources

Background

Regional wholesaler and largest supplier of treated water in the United States serving 19 million people, The Metropolitan Water District of Southern California (The "District") had accumulated 80 TB of data over 30 years of file share use. With 1800 employees, the process to clean up the data was not only a technical feat, but also a daunting organizational project.

A District-wide employee assessment determined the legacy file systems were not meeting the District's needs for sharing information among staff and stakeholders. Staff reported they could not readily find information needed for daily operations and much of the data appeared obsolete.

To improve information access and governance, the District planned to **migrate its file share data to the cloud** and tag it with **searchable metadata.** However, they realized that first the data needed to be understood, cleansed, and better organized.

Setting Up for Success

To ensure a successful cloud migration the District planned a two phase strategy:

1. **Discover:** The District's **more than 200 file shares** were, among other things, an obstacle to productivity. The first step was the daunting task of discovering what data existed on the file shares so then, they could decide was worth keeping.
2. **Decide:** In order to decide the value of maintaining data or discarding it, ActiveNav met with **over 400 stakeholders across 190 sub-business units and teams.** Once the decision to keep the data was made, it was then prepared for migration to the cloud.
3. **Sustain:** Cloud migration and metadata tagging of **40 million files / 80 TB** meant that the District was set up for sustainable records compliance.

At a Glance

ABOUT

- Public Sector Utility
- 1800 Employees
- 17 Major Business Units

CHALLENGES

- 80 TB of content
- more than 5 million folders
- 200+ file shares
- File system data had become an obstacle to productivity

RESULT

- Indexed 40+ million files
- 40% of data was ROT or duplicated
- Restructured to align with business operations

PROCESS

- **Discover** what data existed on the file shares
- **Decide** what had no operational value
- **Sustain** compliance and record management requirements

Originally, the District's data comprised over **40 million files** which included:

- Engineering drawings
- Photos
- Videos
- System files
- All manners of office documents from every department

The District knew it needed automated tools and expert assistance to discover the data and facilitate its cloud migration strategy. It selected the ActiveNav file discovery and management system and engaged professional services for consultation.

> One of the most successful outcomes of this project to date is how quickly peoples' behavior changed. I'd say the vast majority bought in very early on, even people that were nervous about this project.

"I'm confident today that by using and partnering with ActiveNav, the work over the last few years has positioned us well for taking the next step of deploying an ECM system and actually having that technology work as intended because of the foundation built and the prep work ActiveNav allowed us to do," said Steve Gonzales, Information Governance and Enterprise Content Management Manager.

Making Sense of the Data

During phase 1, Discovery, the District's information management specialists team and ActiveNav's consultants indexed over **200 file share** name spaces. The team conducted data review sessions with business units to evaluate suspected ROT data identified by Discovery Center's out-of-the-box ROT rules.

During phase 2, Decide, ActiveNav worked with more than 400 stakeholders from business units to decide on the value of the indexed data. Some business units identified as much as **40% of their file share data as being ROT, redundant, obsolete, and trivial (ROT) data and had no operational value and was no longer subject to record retention.** This ROT would be excluded from migration to the cloud and deleted after final governance review. ActiveNav helped business units **restructure their folders to consolidate them and decide what aligned with** present day operations. The result was a streamlined and logical folder structure containing files cleansed of ROT. Business units enthusiastically participated, saying they had wanted to do file cleanup for years but did not have the tools required.

Gonzales added, "Change management was also a big piece. We've been doing things the same way for a very long time. The technology made it easy so we could truly focus on change management. One of the most successful outcomes of this project to date is how quickly peoples' behavior changed. I'd say the vast majority bought in very early on, even people that were nervous about this project."

40%

40 million files

40% of data was ROT or duplicated
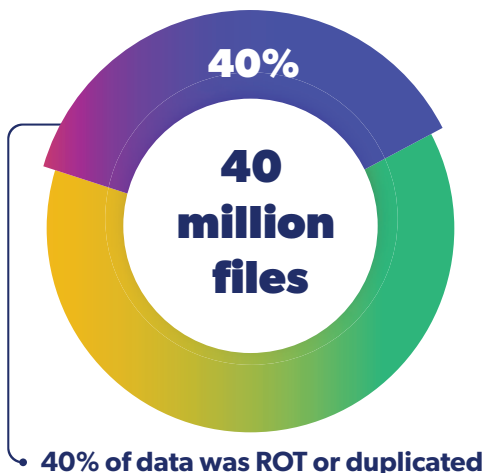
In addition to **data cleansing and restructuring**, ActiveNav's file content analysis capabilities were used in support of other information governance objectives, including:

- **Identifying terminology** used by business units to help update the District's records retention schedule.
- **Identifying** and **categorizing** documents related to specific elements of the District's vast infrastructure such as pumping stations, water treatment plants, dams, hydroelectric plants, reservoirs, pipelines, and canals.
- **Identifying duplicate** and **overlapping folder paths** leading to the same file content.
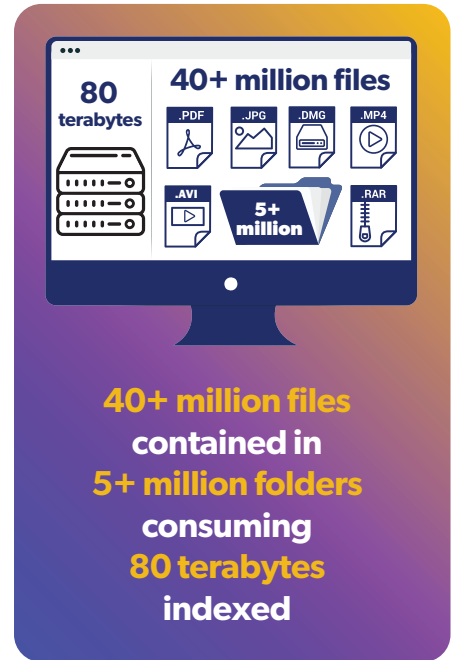
"I want to emphasize that I look at ActiveNav as having a multifaceted approach to our project in particular. It wasn't only the software that got us to where we are today, but the capabilities of the ActiveNav team and how we really, truly depended on that to have success in this area," says Steve Gonzales. "We refer to ActiveNav as our strategic partner and think the combination of their services and its technical solution is really what made us successful.".

Meeting Future Needs

With ActiveNav Governance, the District was able to complete discovery and prepare to take the next step toward cloud migration. The District plans to further utilize ActiveNav in migrating the cleaned-up file shares into its new cloud repository. Once migrated, the District will use ActiveNav's file analysis capabilities to categorize, tag and sustain files.

ActiveNav's combination of services and technical solutions set the District up for the **future compliance sustainability**, which includes remaining compliant and meeting record management requirements. The District recognized how ActiveNav's repeatable process of **discovering**, **deciding** and **sustaining** their information governance helped them get their employees to change their behavior and encouraged them to adopt plans and strategies for the future.

80 terabytes

40+ million files

.PDF .JPG .DMG .MP4

.AVI **5+ million** .RAR

40+ million files contained in 5+ million folders consuming 80 terabytes indexed

The Bottom Line

> Again, without ActiveNav I don't know that our organization would be ready today to take those next steps. Thankfully, it was a very enjoyable project because of those outcomes, and certainly the highlight of my career that, again, I was able to enjoy thanks to that partnership and the outcomes and success that we accomplished together.
>
> Steve Gonzales, Information Governance and Enterprise Content Management Manager at The Metropolitan Water District of Southern California.

GET STARTED TODAY. SEE ACTIVENAV IN ACTION. REQUEST YOUR DEMO.