# ACTIVENAV

Purview for Data Discovery Use Cases? How's That Going for You?

Through Purview, Microsoft is promising the world for information governance. However, for organizations expecting that to be delivered, there are important gaps to be aware of. Further, with such a broad scope, any successful Purview implementation requires significant investments. Beyond E5 seat licensing, consulting for implementation and internal technical expertise for sustainment, perhaps the biggest investment challenge is time. Specifically, the time required to receive promised roadmap updates and achieve required capabilities.

If these investments and delays are okay then, we're genuinely happy for you, and you need read no further. However, for those who want to get things done, this paper describes why acquiring more focused solutions for data discovery should be a key consideration by any organization exploring Purview. It then, in a series of frequently-asked-questions (FAQs), explores the less obvious real-world issues with using Purview for data discovery and related use cases.
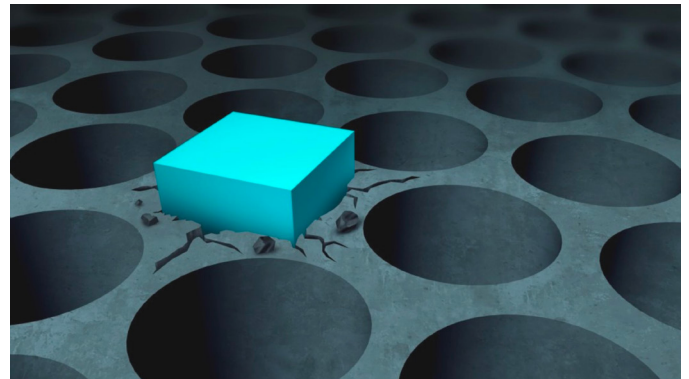
What's the Problem?

With organizations struggling under the weight of ever-growing unstructured data assets and compliance demands on the rise, data discovery should be considered a foundational step in many, if not all, information governance projects. However, despite advances in key technologies, the discovery of unstructured data remains a hard problem to solve. From file shares to mail servers, chat to collaboration platforms, document management to systems of records, the fact is that data discovery is important because the data within lacks good context and users have failed to organize or maintain it. This has resulted in unstructured data repositories becoming a tangled mess of records and other content which can be all but impossible to resolve without the right technology to enable dedicated records or governance teams.

Despite the hype, modern search and classification techniques and services struggle to efficiently handle large data volumes. What's more, organizations understandably set a high bar for accuracy; this means that data discovery solutions need to effectively combine a range of additional capabilities to make data discovery successful and sustainable for key governance use cases such as data inventory and mapping, data cleanup and sensitive data identification and minimization. The greater the scale of the project, the older, more chaotic the data, the more this is the case.

The remainder of this paper and accompanying FAQ section goes into detail on the required capabilities for these use cases, but we'd summarize them as follows:

- Consolidated, aggregated discovery and reporting across a variety of repositories, both on premise and in the cloud.
- Reporting at scale well beyond 100s of terabytes.
- Data visualization, usable by non-technical staff, which enables efficient decision making on the right data.
- The ability to handle multiple criteria (such as aging data AND sensitivity) for efficient data disposition.
- Effective navigation around, through and deep into target data, while retaining context.
- Specific workflow support for quality control and review, which scales across an organization.

Americas
Reston, VA, USA
+1 571 375 2780

EMEA
Winchester, UK
+44 01962 454401

APAC
Melbourne, Australia
+61 3 9982 4543

Contact sales@activenav.com
Tweet @ActiveNav
Visit ACTIVENAV.COM

So, for organizations who want to get started in data discovery and be on their way to a future where data is better governed and understood, a focused solution that addresses these features is required. By contrast, selecting do-it-all solutions, such as Purview, from vendors with little experience in information governance use cases can be expected to lead to project delays and loss of stakeholder support as valuable outcomes become all but impossible to achieve in any relevant timeframe.

Why Purview Misses the Mark

While the people behind Purview are smart, the problem is that they are attempting to build a very broad capability of which data discovery is only a small part. As a result, there are many gaps not only in the feature set but, more importantly, in the ways those features are drawn together to build a working capability. This is especially the case for data discovery use cases because Purview is implemented in the M365 ecosystem, where content is intended to spend its entire lifecycle. This gives rise to key limitations for any such use case:

- Any repository outside of M365 is poorly handled, with a limited feature set. For example, file shares are handled through a separate application with limited reporting or integration with other repositories.
- File formats beyond core office formats are unsupported, meaning the contents of archive formats or legacy formats remain opaque.
- Purview's use case tools have been developed as separate capabilities. This means, for example, that retention and data sensitivity cannot be considered in the same context which is highly inefficient in any organizational-scale implementation.
- Regardless of the content source, Purview fails to report adequately, at scale. Further, it fails to report broadly across diverse connected repositories. Such capabilities are fundamental to a consistent and sustainable governance program.
- Purview's reporting lacks detail. That is, it fails to provide the range of supporting metadata necessary for the use to take disposal decisions with unstructured data where storage context is of paramount importance.

The importance of the above limitations becomes clear with experience, experience that most organizations don't get the chance to grow, much less apply to technology procurement decisions. This leads to procurement decisions based upon a checklist of features and functions where Purview's breadth is both impressive and alluring. However, lacking experience, customers consistently learn that the way features and functions are put together (or, rather, the gaps between them) is critical to harnessing the potential of data discovery technologies. Having fallen into those gaps, customers who reach out for support from Microsoft will find that their teams are learning too. This means the customer's users become an extension of the Microsoft quality team rather than productive in the intended project.

Finally, Microsoft Purview is designed very much from the perspective of IT teams. By contrast, data discovery for governance necessitates that governance teams, records teams, analysts, and business subject matter experts comprise the core user base, since only these groups can make data disposition decisions. This compounds the problems described above. In short, data discovery is for the business, not for IT.

Americas
Reston, VA, USA
+1 571 375 2780

EMEA
Winchester, UK
+44 01962 454401

APAC
Melbourne, Australia
+61 3 9982 4543

Contact sales@activenav.com
Tweet @ActiveNav
Visit ACTIVENAV.COM

ActiveNav Cloud, a Specialist Data Discovery Service

ActiveNav has evolved and grown its experience through the school of hard knocks with technology, product and consulting teams engaged with heavy data discovery projects across a broad range of use cases. These projects are exactly what Purview does poorly. In contrast, ActiveNav Cloud has been built from the ground up to capture decades of experience in a single service designed specifically for data discovery across the broadest set of unstructured data repositories. It doesn't require any infrastructure, nor is it attempting to solve all the problems Purview claims. Also, it doesn't make any assumptions about the target data or the repository it's stored in.

Our customers don't want to hang around. They've built the business case, they've seen the demo and they have their stakeholders aligned. They know that if they don't produce results quickly their support and buy-in will wither and the project will stall.

ActiveNav Cloud data discovery runs fast, is highly configurable and results are available almost immediately presented with all of the necessary context. Furthermore, once the results are available, there is no need to go anywhere else to take disposition decisions. This means bulk data handling and cleanup is efficient and defensible. All of this is because we've been through the objections and roadblocks and implemented the compensations to account for them.

Finally, we know that Microsoft is here to stay, and we know that our customers are often dealing with their data problems as part of moving to Microsoft once and for all. For that reason, ActiveNav Cloud integrates with Microsoft and Purview to make that transition as easy as possible. Simply put, when dealing with data discovery use cases, ActiveNav Cloud is a tool designed for the job that will enable you to get started fast, grow momentum, and get a grip of your entire unstructured data estate without significant investment or startup costs that so frequently kill information governance business cases.

In Summary

As we noted earlier, a feature-by-feature checklist doesn't tell the story and so the following captures an alternative, real-world, summary for ActiveNav Cloud and Purview, side-by-side.

Capability	ActiveNav Cloud	Microsoft Purview
Readily discovers and reports on data outside M365 in the same way as that inside M365	●	○
Flexible configuration and support for broad use cases with unlimited search terms	●	◑
Addresses bulk data discovery and reporting at scale, out of the box	●	◑
Handles data in on-premise or cloud file shares, or elsewhere	●	○
Provides scalable reporting and visualization through a single pane of glass across all repositories	●	◑
Efficient and consistent reporting, usable by non-IT staff, with drill down and context across all repositories	●	○
Readily extensible and configurable for a broad range of data discovery use cases	●	◑

Frequently Asked Questions

We're committed to Microsoft Purview, why invest in a tool that does the same thing?

Purview has been designed as part of the wider M365 ecosystem. This brings with it several important disadvantages. First, and perhaps most importantly, it supports only M365-hosted content and formats (such as file shares or specialist document management systems), effectively excluding other content from effective governance. Second, Purview's search configuration has important limitations (such as the scope and number of keywords) which limit the range of use cases it can practically support. Third, Purview does not prioritize data discovery, which means it fails to provide insights necessary to enable effective disposal decisions to be made at scale, a problem that causes many data discovery implementations to fail.

Investing in ActiveNav Cloud (which has been designed specifically for effective and flexible data discovery) will ensure you get a complete picture of your data and that your governance team, alongside data owners, can actually make the decisions necessary to mitigate data risk.

Purview has trainable classifiers, what else could I need?

Purview provides a range of different classifiers, some of which use training models for configuration. Each type of classifier has its own strengths and weaknesses across different use cases, with trainable classifiers having most utility for a small number of focused data types, or specific use cases, with the disadvantage of being opaque for users trying to understand their behaviors. By contrast, other models are faster to configure and more transparent in use which makes them more practical for use in many data discovery situations. Combined with other limitations, this means that Purview is tough to configure for the use cases that are important to your organization.

Purview can connect to file shares, why not do that?

The application provided by Purview for handling data discovery in file shares is a standalone tool which works almost entirely in isolation from the M365 environment or, indeed, other repositories. What's more, its reporting capabilities are limited both in terms of the use cases supported and the scale over which they can be used. This means that deploying Purview's tools for large file share data discovery will require a significant amount of additional technical resources to be consumed with limited returns from the project. Perhaps most importantly, it will isolate data owners and users from the data which brings increased risk in terms of both business buy-in and the accuracy of results. By contrast, ActiveNav Cloud is designed to enable teams to focus on the data, rather than the technology, and address the full range of data discovery use cases consistently against all unstructured data regardless of its location.

About Us

We're data experts, and our North Star is Zero Dark Data. We believe that all organizations should be aiming for a state of Zero Dark Data so that they can act as good stewards of that data to minimize their cyber risk surface area, protect the interests of their customers, their staff and their other stakeholders. We've been working continuously with unstructured data in the wild for well over a decade and we think the market deserves data discovery products that just work. As a result we are trusted by leading companies and government agencies to help them understand and control their data assets to drive regulatory compliance, reduce the cost of data ownership and improve data quality.

Aside from our fascination with Dark Data, we're engineers, designers, runners, gardeners, chefs, photographers, bikers, parents, skiers, cosplayers, hikers, gamers, travelers, mountain climbers, friends, race car drivers, readers, volunteers, and car enthusiasts. We value loyalty, accountability, and communication and care deeply about creating a sustainable future, supporting charitable causes as proud members of Pledge 1%.

→ 15 Years of Experience
→ 6 Continents and 28 Countries
→ 15+ Billion Files Discovered
→ 30K+ hrs Customer Deployments
→ 300+ Customers and Counting

Americas
Reston, VA, USA
+1 571 375 2780

EMEA
Winchester, UK
+44 01962 454401

APAC
Melbourne, Australia
+61 3 9982 4543

Contact sales@activenav.com
Tweet @ActiveNav
Visit ACTIVENAV.COM