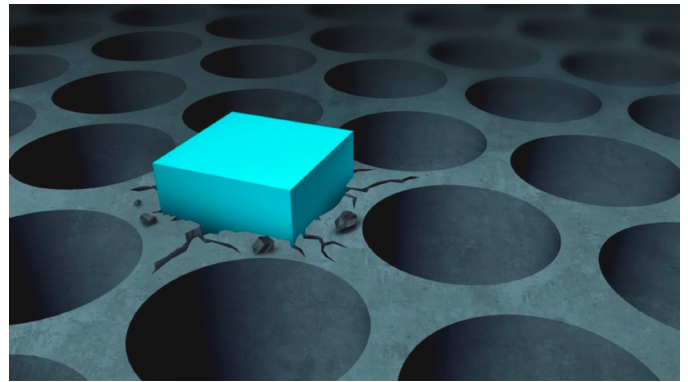


Purview for Federal Data Discovery and Cleanup

Use Cases? How's That Going for You?

Through Purview, Microsoft is promising the world for information governance. However, for federal agencies 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 then that required to actually achieve a capability.

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 agencies exploring Purview. It then answers a series of frequently-asked-questions (FAQs), exploring the less obvious, real-world issues with using Purview for data discovery, cleanup and related use cases.



What's the Problem?

With agencies 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, governance projects. However, despite advances in key capabilities, 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, agencies 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 meaningful for key governance use cases such as data cleanup, records and sensitive data identification or security labelling. 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, but we'd summarize them as follows:

- Aggregated discovery and reporting across a variety of repositories, both on premise and in the cloud.
- Handling of data at scales 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 (for example, aging data AND sensitivity) for efficient data disposition.
- Effective navigation around, through and deep into target data while retaining context.
- Specific support for difficult clean up use cases such as duplication removal and records identification.



So, for agencies 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 to meet their needs. 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 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. 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 project-based data discovery use cases because Purview is implemented in the M365 ecosystem, where content is intended to run 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 retention and data sensitivity cannot be considered in the same context which, while potentially acceptable during steady state use, is highly inefficient in any large-scale project.
- Regardless of the content source, Purview fails to report adequately, at scale. Further, it fails to report broadly across diverse repositories. Such capabilities are fundamental in the early stages of any discovery project for scoping and orienting data sets.
- 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. It's easy to make procurement decisions based upon a checklist of features and functions when 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 in data discovery projects. Having fallen into those gaps, customers who reach out for support from Microsoft will find that they 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. However, in data discovery, governance teams, records teams, analysts and business subject matter experts are 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.



ActiveNav's DC Project Suite

DC Project Suite has been evolved through the school of hard knocks by teams engaged heavy data cleanup and enrichment projects. These projects are exactly what Purview does poorly. In contrast, DC Project Suite captures decades of experience in a single toolset designed specifically for the data cleanup and enrichment, especially for legal data at scale. It doesn't require extensive 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.

DC Project Suite's discovery runs fast, is highly configurable and visual tools help test and validate rules and policies by which the business experts that understand them. Results are available very quickly and they present all of the necessary context in the right place. Further, once the results are available, there is no need to go anywhere else to take action, regardless of the use case. This means bulk data handling and cleanup is efficient and is all in one place. All of this is because we've been through the objections and roadblocks and implemented the compensations to get around 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, DC Project Suite integrates with Microsoft and Purview to make that transition as easy as possible. Simply put, when dealing with data discovery use cases, DC Project Suite is a tool designed for the job that will enable you to get started fast, grow momentum and roll into the M365 stack with your data in the best shape it's ever been in.

In Summary

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

Capability	DC Project Suite	Microsoft Purview
Readily discovers and reports on data outside M365 in the same way as that inside M365	●	○
Addresses bulk data projects at scale, out of the box	●	◐
Handles legacy data in file shares or elsewhere	●	○
Provides flexible reporting and visualization for cleanup, sensitive and records use cases	●	◐
Efficient and consistent reporting, usable by non-IT staff, across with drill down and context across all repositories	●	○
Readily extensible and customizable for records, CUI and other classification for agency-specific needs	●	◑



FAQ: Why DC Project Suite When You Have Purview?

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

Purview simply isn't built for large data discovery and cleanup projects. It's been designed to provide governance features as part of the M365 ecosystem during the normal information lifecycle and not to address big batches of data as part of a change or migration project. For that reason, investing in DC Project Suite (which has been designed for large unstructured data projects) will enable your core Purview team to get on with managing the M365 environment, without distraction, while a focused project team can get started fast and use purpose-built tools to clean up and prepare non-M365 data to be in the best possible shape for its migration into the new environment.

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 high value, focused, data types or specific use cases with the disadvantage of being opaque for users trying to understand their behaviors. By contrast, for bulk data cleanup and migration projects, other models are faster to configure and more transparent in use which makes them more practical for use in such situations.

Microsoft provides out of the box templates for CUI and NIST SP 800-171. Why do I need DC Project Suite?

The difference lies not in the rules and classification templates but in the design of the tools within which those rules are deployed. The same templates are available in DC Project Suite but, in contrast to Purview, DC Project Suite provides extensive support for transparent rules customization to meet agency-specific needs while shining in its ability to handle large-scale data clean up and/or preparation projects. That means, regardless of the required rules and classifiers, you should use DC Project Suite for any such bulk data project, as described above.

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 projects 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 project which means project risk is increased in terms of both business buy in and the accuracy of results. By contrast, DC Project Suite is designed to enable teams to focus on the data, rather than the technology, and address the full range of clean up use cases (from sensitive data, through ROT to duplication and classification) consistently against all unstructured data irrespective of its location. It connects the business to its data to promote business buy in and integrates with M365 and Purview so that prepared data can be migrated quickly and efficiently.



Purview does records labeling for retention, why not just build around that?

In short, it's really about a combination of all of the above responses. The issue is not really about retention classification and labeling but more about how the toolset deploys those capabilities in bulk data projects. Simply put, Purview isn't designed for such projects. By contrast, DC Project Suite is designed for that very purpose and it is integrated with Purview so that its records classification results can be determined in advance and the appropriate retention policy can be applied as soon as the data arrives in M365. This gives you the best of both worlds; data prepared and cleaned up quickly ready for M365 and M365 is used most effectively as soon as data arrives. Finally, to help get moving fast, DC Project Suite provides a pre-configured implementation of NARA's General Records Schedules ready for deployment and adaptation to each specific agency's needs.

Purview provides cool dashboards; why does DC Project Suite's reporting seem limited?

DC Project Suite is designed to do a job; that is, discovery, clean up and classification of large data sets as part of a wide range of governance projects. This means that its reporting is focused on providing maximum utility for specific project-based use cases, from ROT removal through duplication clean up to sensitive data and record labelling. In addition, experience shows that data discovery projects are more efficient when reporting can handle multiple criteria (such as aging data AND sensitivity), something hard to achieve with Purview. Finally, these reports work across all connected repositories and can be configured to match the shape of the agencies while Purview operates in silos and provides more of an overview than focused reporting.

How can you support M-23-07 beyond what I'm doing with Purview?

Any agency seeking M-23-07 compliance must address the breadth of its electronic data assets, regardless of their location. This, and other requirements, demand much from data discovery technologies and, specifically with respect to Purview when addressing data outside of the M365 environment. Simply put, deploying records classifiers outside of the M365 is challenging with Purview and that is compounded by its limited support for the design, customization and modelling of classifiers for agencies' specific records schedules. Further, Purview does not readily support the creation of data packages which support NARA's metadata standards for permanent records transfer. In combination Purview's other shortcomings for dealing with bulk datasets (described above), M-23-07 compliance is made much easier using DC Project Suite alongside (and integrated with) Microsoft Purview.

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