Duquesne University Finds and Tags Digital Images 99% Faster with Komprise and Amazon Rekognition

Overview:

Duquesne University is a private Catholic research university in Pittsburgh, Pennsylvania, founded in 1878. The university is ranked in the prestigious Princeton Review and has achieved several rankings in the US News & World Report annual college rankings including a "Best Value School" for 2024.

Duquesne's IT organization was seeking insights into its storage environment to help them manage quickly growing unstructured data volumes on its on-premises storage arrays, which was creating a need to purchase more capacity at a faster rate all the time. Duquesne's storage systems administrator Matt Madill wanted to know how much of the data was cold or rarely used so he could right-place it for dramatic cost savings. This requirement led them to Komprise Intelligent Data Management. After using the product to analyze its silos of unstructured data and **tier cold data to the cloud**, a new use case emerged for AI and image recognition.

Key Stats

Industry: Higher Education

Storage: Various NAS, AWS, Wasabi

Use cases: Smart Data Workflows, Data Tiering, Deep Analytics

Results: Komprise + AWS Rekognition delivered 99% reduction in time for image search; repeatable use cases for new AI data workflow automation

1.888.955.0290 | komprise.com

Komprise and Smart Data Workflows with AI

Duquesne's library archive department regularly fulfills requests for files from different University departments. For instance, the head of fundraising might want to see all the images of celebrity donors for a new campaign. But given the millions of digital images spread across different offices, it was an extremely manual, time-consuming effort. Some images have descriptive file names, while others do not. The University needed a way to look beyond the metadata, inside the file contents, to identify the right images.

A new request came up to find images of U.S. President John F Kennedy and a famous campus building for the marketing department. The archivists identified about 10,000 files of interest. Assuming each file would require at least two minutes to manually inspect, they estimated it would take at least 20,000 minutes or 333 hours to fully review and record the results. A similar project the previous summer took most of the time of one of the archivists. When Madill found out about Komprise Smart Data Workflows Manager, he considered there might be a better way.

- **Smart Data Workflow Manager** has a simple, point-and-click interface so users can create automated workflows to find the right data across storage assets, tag and enrich the data for improved classification, and integrate the workflow with AI and ML tools.

- Users don't need any specialized coding or AI skills to set up a workflow; an intuitive wizard takes them through all the steps to select the data parameters and configure the AI service.

- The Komprise **Global File Index** and Smart Data Workflow Manager together reduce the time it takes to find, enrich and move the right unstructured data to AI by up to 80%.

- Komprise can also integrate with **Amazon Rekognition** to execute a Smart Data Workflow that automatically finds and tag images with relevant metadata for indexing.

Duquesne deployed Komprise Smart Data Workflow Manager with Amazon Rekognition to automate the process for two use cases, driven by the library archive team:

1. For a digital exhibit, the team needed to find images of JFK visiting the campus across more than 10 terabytes of storage on the university's cloud services.

2. Image recognition for specific buildings, which requires creating custom labels.

Search	→	Run AI	→	Tag

- Komprise knows where the University Archives images are located
- Search for Archival images only

- Komprise calls Amazon "RecognizeCelebrities"

- Komprise tags University Archive images with JFK and other celebrity tags;
- We can now search for JFK in Komprise to inspect and override false positives

Komprise automates the workflow of curating data by feeding Amazon Rekognition using both standard and custom LLMs. Users begin with a Komprise Deep Analytics query locating potential files with JFK images across multiple NAS systems. The files identified by the query were sent to Amazon Rekognition for scanning and the resulting labels were mapped to tags in the Komprise Global File Index.
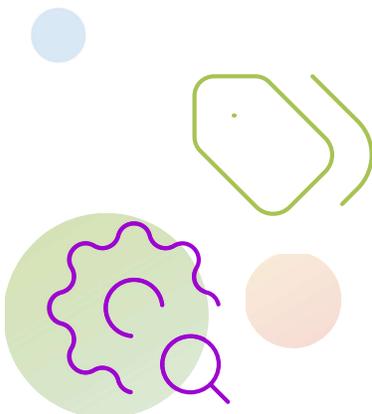
The workflow included a schedule to continuously scan new or modified potential files. The process was not only much faster and more accurate with this automation but resulted in longer-term cost efficiencies. The Duquesne team also plans to use the Komprise Smart Data Workflow Manager for PII detection.
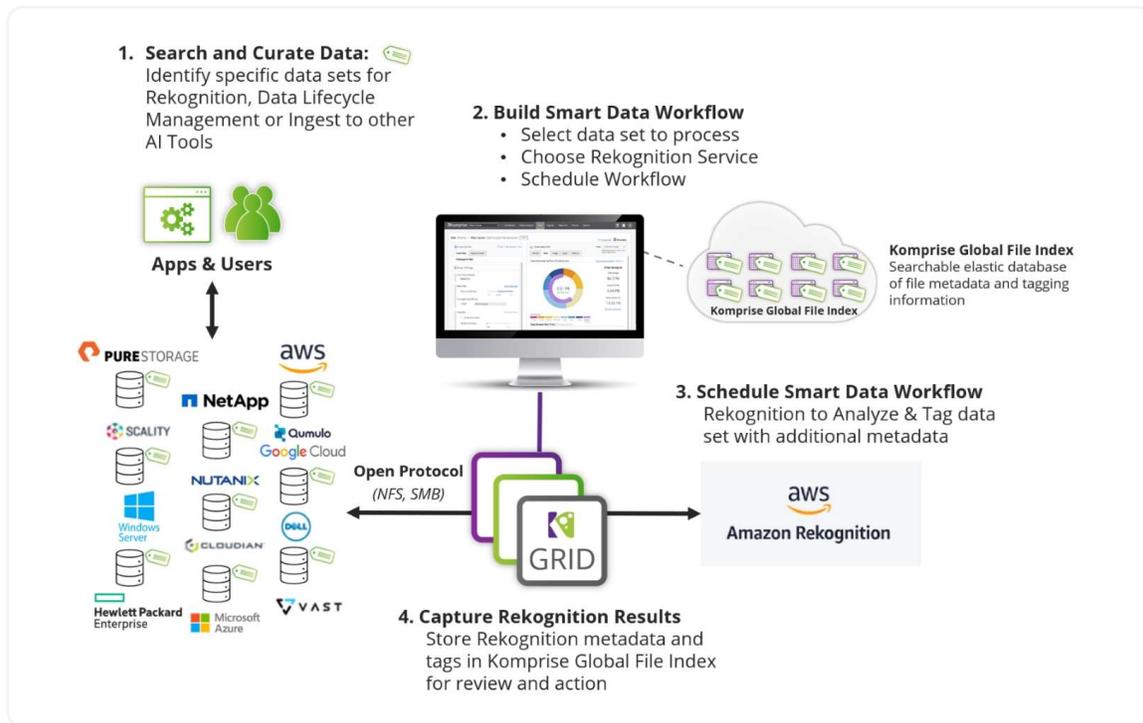
"

"Our digital collections are growing in leaps and bounds but budgets stay flat," said Rob Behary, Head of Systems and Scholarly Communications, Gumberg Library at Duquesne University. "AI is still new but has tremendous potential. With Komprise we're able to improve efficiency with a systematic workflow to index data, run AI and tag data."

Rob Behary,
Head of Systems and Scholarly Communications,
Gumberg Library at Duquesne University.

1. **Search and Curate Data:** Identify specific data sets for Rekognition, Data Lifecycle Management or Ingest to other AI Tools

Apps & Users

2. **Build Smart Data Workflow**
   - Select data set to process
   - Choose Rekognition Service
   - Schedule Workflow

Komprise Global File Index Searchable elastic database of file metadata and tagging information

3. **Schedule Smart Data Workflow** Rekognition to Analyze & Tag data set with additional metadata

Open Protocol (NFS, SMB)

GRID

Amazon Rekognition

4. **Capture Rekognition Results** Store Rekognition metadata and tags in Komprise Global File Index for review and action

PURESTORAGE · aws · NetApp · SCALITY · Qumulo · Google Cloud · NUTANIX · Windows Server · DELL · CLOUDIAN · Hewlett Packard Enterprise · Microsoft Azure · VAST

Benefits of Komprise Intelligent Data Management

Komprise, an unstructured data management and mobility SaaS, brings numerous capabilities to Duquesne from analytics of file and object data across all storage; transparent tiering of data to secondary storage for cost savings; Deep Analytics capabilities to run ad hoc queries and Smart Data Workflows.

Duquesne's top benefits with Komprise include:

Rapid fulfillment of image request: The entire Smart Data Workflows process from specifying the data and configuring the AI service to the final output took 1 hour and 53 minutes. In this scenario, the time spent went from 20,000 minutes of time to search across directories manually to 113 minutes — a 99% reduction in time and a massive productivity gain for the library archive team.

Retain AI results for efficiency: Komprise keeps the tags in its highly-scalable Global File Index so that users can search for files in Komprise without re-running the AI process. This is an important detail, since third-party AI services charge by the job.
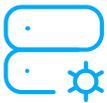
Policy-based automation: Duquesne can configure a policy for continual workflow processing so that the results are always up to date. Komprise allows policy-based automation on any number of data management tasks, such as copy, confine, migrate and/or tier.

Practical, repeatable use cases with AI: The Komprise Intelligent Data Management integration with third-party AI tools for automated data workflows is the perfect example of a discrete use case for AI with demonstrable benefits and low risks. Beyond image search and inspection, this is a repeatable use case which can also identify files to be deleted or archived based on the content—an opportunity for significant cost savings in data storage and backups. In another example, a scheduled policy with Amazon Rekognition can scan newly arriving files, tag any with inappropriate content, then execute a scheduled Plan Action to automatically move the files to a restricted-access storage area for review.

No coding or AI expertise required: Komprise Smart Data Workflows Manager was designed for use by any authorized user. A wizard walks users through the steps of searching for the right data set, configuring and tuning the AI service, defining the tags and selecting how frequently the workflow should run. Users can monitor hundreds of workflows from a single interface, even when leveraging AI services from different clouds.

Data governance and auditing: Komprise delivers data governance with auditing by maintaining logs of details such as what data is fed, when, and to which service. This capability is crucial for protecting sensitive corporate data from being ingested by external LLMs and exposed to users outside of the organization.

> " "Smart Data Workflows is a natural extension of our use of Komprise beyond cold data storage cost optimization," says Matt Madill, Storage Systems Administrator at Duquesne University. "Using AI, Komprise is giving our departmental users a powerful way to search based on both metadata and file contents."

Matt Madill,
Storage Systems Administrator at Duquesne University.

komprise