Diskover

SOLUTION BRIEF | HEALTHCARE

From first scan to long-term archive, every file accounted for.

One enriched metadata catalog for clinical, imaging, and research data across NAS, object, cloud, and archive.

90% of the data organizations generate is unstructured, per IDC

$542K reclaimed within days of the first index at one cancer center

1.4B files indexed in two hours, ~200K files per second

At a glance

KEY CHALLENGES

Hospitals hold petabytes of clinical records and imaging across archives, shared drives, and clinical apps, with no shared inventory. Likely PHI has no clear owner, scanners leave temporary files on premium storage, and HIPAA and GDPR retention runs on spreadsheets.

THE SOLUTION

Diskover indexes every file into one metadata catalog, with the data staying in place. Teams locate likely PHI by metadata such as path, file type, and owner, clear temporary imaging output by policy, keep an audit trail, and send curated datasets to AI pipelines.

WHY IT MATTERS

Audit requests are answered in minutes instead of days. Premium capacity comes back from temporary and duplicate imaging files, so the next purchase is smaller or later, and clinicians and researchers keep the capacity they need.

Why Diskover

ONE CATALOG FROM SCANNER TO ARCHIVE

Index every file across clinical NAS, imaging archives, research storage, object, cloud, and tape into one searchable view by fields such as department, project, format, and age, so one search covers active studies, slide images, and archive, with data staying in place.

LIKELY PHI, LOCATED BY METADATA

Find the files most likely to hold PHI by metadata such as location, file type, owner, and age, plus file headers when plugins enrich them. Tag them, report on them, and keep them on the right tier. The approach is fast, non-invasive, and built for petabyte scale.

IMAGING CLEANUP BY POLICY

Slide scanners, viewers, and image management platforms leave temporary, transient, and test files on primary storage. Diskover identifies them by metadata such as extension, path, and age, then policy clears or tiers them on schedule, so premium capacity comes back without scripts or manual cleanup.

RETENTION YOU CAN SHOW

Tag the records that HIPAA, GDPR, or your own retention policy requires you to keep, protect them from cleanup policies, and run removal with a preview and a full log of what went, when, and by whom, ready for an audit.

CLEAN DATA FOR RESEARCH AND AI

Curate imaging and clinical datasets by fields such as project, age, or tag, leave out files flagged as likely PHI, and deliver the metadata to research and AI pipelines through the API, with source files staying under the hospital's control.

ASK IN PLAIN LANGUAGE

Ask the catalog, "Which slide images older than five years are still on premium storage?" and get the matching files back with a recommended action, such as tagging, tiering, or packaging for a researcher. AI-assisted, human-approved.

About Diskover

Diskover helps enterprises structure unstructured data across cloud, on-prem, and edge environments, making it easier to search, govern, optimize, and deliver file data for compliance, analytics, and AI-ready workflows.

Contact

diskoverdata.com
sales@diskoverdata.com
(800) 560-5853
San Jose, CA

Diskover

SOLUTION BRIEF

How it works

1 CONNECT
STORAGE OR IT TEAM

Point Diskover at clinical NAS, imaging archives, research storage, object, cloud, and tape. Indexing runs in place, so images and records stay where clinical systems expect them.

Example: the pathology share is cataloged where it sits. Files are not copied or moved.

2 CATALOG
AUTOMATION

Every file becomes a searchable record with rich metadata, such as name, path, and owner, enriched with path tokens, file headers from plugins, and classification tags.

Example: a slide image now shows its owner, last access date, and department tag.

3 CURATE
ANY AUTHORIZED TEAM

Tag by fields such as department, study, and retention class. Build saved searches for what teams ask weekly, such as scanner temp files on premium storage, or likely PHI outside approved locations.

Example: a researcher or clinician saves "old slide images" as a search.

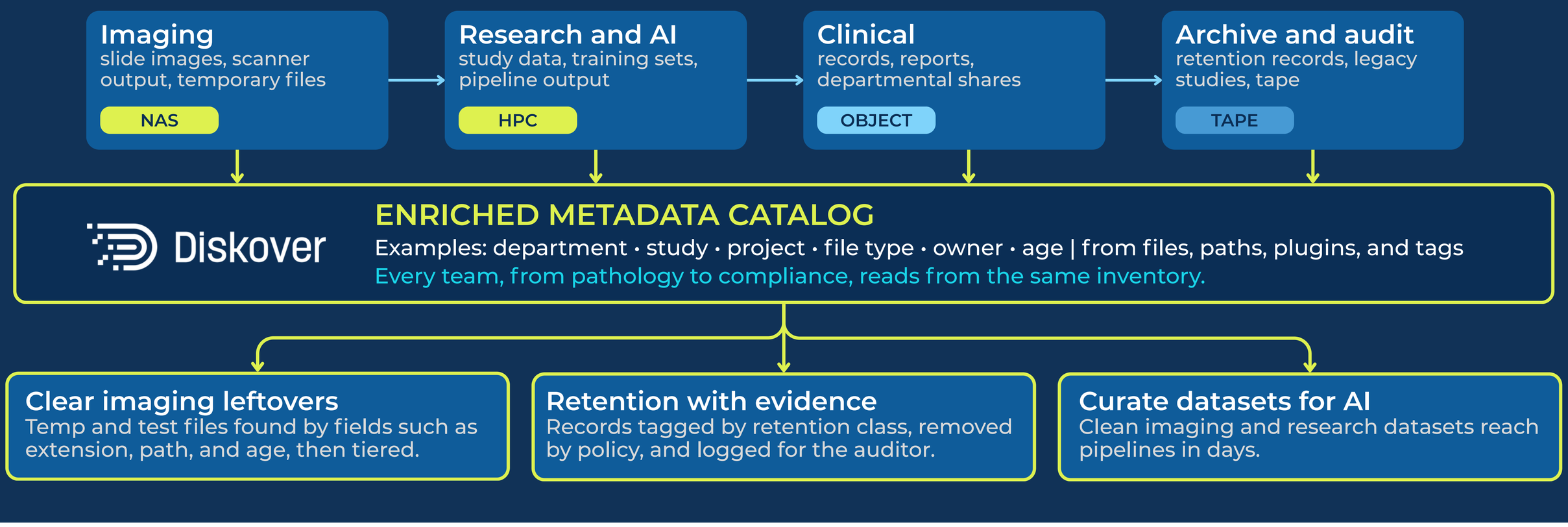
4 ACTIVATE
BY POLICY

Run tiering, removal, and retention by policy with a full audit trail, produce reports on demand, and deliver curated metadata to research, analytics, and AI pipelines through the API.

Result: at one cancer center, 1 PB of files over five years old was flagged for tiering.

WORKS ACROSS Cloud · On-prem · Hybrid · NAS · SAN · Object · Tape · Archive · Data lakes · Lakehouses · AI/BI/ML

Every scan, study, and record in one view.



Key outcomes

AUDITS ANSWERED FAST

Compliance and legal teams find likely PHI and retention records across every system in one search, then export the evidence with fields such as owner, location, and retention status.

STORAGE THAT STAYS CLEAN

Scanner temp files, duplicate images, and abandoned test output move off premium tiers by policy, so capacity comes back on its own and the next storage purchase is smaller or later.

RESEARCH AND AI READY

Curated imaging and clinical datasets reach research and AI pipelines in days instead of weeks, with likely-PHI files flagged before they move.

GET STARTED

Ready to see it on your estate?

Start with a 30-day proof of concept on your own data, and recover your investment in 3 to 6 months.

About Diskover

Diskover helps enterprises structure unstructured data across cloud, on-prem, and edge environments, making it easier to search, govern, optimize, and deliver file data for compliance, analytics, and AI-ready workflows.

Contact

diskoverdata.com
sales@diskoverdata.com
(800) 560-5853
San Jose, CA

Revised November 2026