

OpenText Knowledge Discovery Collection and Enrichment

Quick view

- Built-in AI for intelligent data discovery*
- Easy access to more than 160 data repositories
- Comprehensive set of out-of-the-box data connectors
- Bi-directional data connectors for content synchronization and governance**
- Embedded ETL framework
- Mature and repeatable platform API
- Extensible SDK for custom data connector development
- Big data scalability
- Straightforward OEM integration with comprehensive documentation
- Resource efficient—no heavy-weight servers required

* Applies to web connector

** Applies to select connectors

Connectors are a core component of the OpenText™ Knowledge Discovery (IDOL) platform, providing a flexible, precise approach to your data management needs. They uniquely aggregate diverse forms of structured, semi-structured, and unstructured data using repeatable and standardized workflows, while managing diverse data in place to minimize duplication requirements, storage costs, and handoff risks. Connecting to a wide range of disparate content repositories and supporting more than 2,200 file formats, connectors offer a simple, fast, proven way to break down data silos for enterprise search and data analytics.

Product highlights

Making the connection to your data shouldn't be complicated. With our advanced suite of connectors, we help you solve critical information connectivity challenges so you can focus on what's most important—running your business and not worrying about your systems.

Connect to your information

Today's organizations are faced with an explosion of unstructured data coming from emails, images, audio, video, online sources, and more. This fragmented and siloed information is driving the need to easily connect to data across disparate content repositories, such as Microsoft 365, Teams, SharePoint, Google Workspace, OpenText Content Management, and others both on premise and cloud-based platforms. With out-of-the-box data connectors, you're able to perform data aggregation and manage across hundreds of content repositories and govern your data using standardized APIs.

While you can request your own in-house data connectors to integrate with your systems, the resources and time required to maintain and keep them going can be costly, while also impacting productivity. Additionally, content repositories are often live systems and interaction with them may compromise day-to-day operations. Your connectors need to interact with repositories in a non-disruptive manner using schedules and incremental processing without being obtrusive.

Increased market requirements are driving the need for connector functionality that goes beyond data synchronization. Diverse unstructured data sets require new methods for governing data via advanced, data-driven workflows, such as content synchronization, content disposition, content syndication, content legal/regulatory compliance via collection and legal hold workflows.

Help increase your analytical potential

With a comprehensive set of data connectors, including those for social media and big data (e.g., Facebook, X, Hadoop, and others), it doesn't matter where your data resides or what format it's in, OpenText Knowledge Discovery can help you govern it. The connectors extract the repository-level metadata and full content of a given target file. Whether for a local or remote content repository, it imports the data into an IDX or an XML file, and then indexes the data into the OpenText Knowledge Discovery Server for information processing. This allows you to perform a much richer data analysis of your content and gain competitive advantage. You can use OpenText Knowledge Discovery to search and categorize millions of tweets, videos, documents, call recordings, and images, while seamlessly delivering real-time, advanced statistical analysis of the topics, demographics, buying behavior, sentiment, geotags, and hyperlink references extracted from that human data.

Not using an OpenText-powered application? Not a problem. OpenText Knowledge Discovery's Connectors give the flexibility to index data into and from third-party applications. *An extensive list is available at the URL included on the last page of this Data Sheet.*

Do more with OpenText Knowledge Discovery's Connectors

Go beyond the basics with our advanced connectors that can enrich repositories with new data or metadata or move information in and out of systems to meet business needs or adhere to specific market requirements. We help you to effectively govern your information and initiate advanced data synchronization, access, and viewing functionality. Standard features include:

- **Sync**—Preserve the CRUD operations done at the data source and have those operations reflected in the generated index.
- **View**—Enable the process of viewing data within the source.
- **Collect**—Enable the connector to "collect" data from source to make it unalterable.
- **Security**—All security protocols are respected and ACLs are transferred as part of the data load.

Advanced features include:

- **Hold**—Mark a data asset as immutable
- **Release**—Mark a HOLD document as no longer HOLD
- **Delete**—Delete a document at source
- **Insert**—Insert a document into the source

With our SharePoint Connector, for example, you're able to index and enhance metadata associated with SharePoint documents, sites, and users, including information stored in list columns, user profiles, and more. Plug in any number of OpenText Knowledge Discovery modules (e.g., categorization, entity extraction, etc.) in the connector layer to further structure the ingested data during the importing process.

OpenText Knowledge Discovery Connectors can help you respond in critical situations. You can initiate legal holds on electronically stored information in SharePoint, OpenText Documentum Content Management, Exchange, and more. The mature connector framework enables governance, access, and visibility across your entire enterprise corpus, allowing for unprecedented control and leverage of your information assets, and to support search that is fully compliant with the Federal Rules of Civil Procedure (FRCP).

Another way OpenText Knowledge Discovery Connectors can help is when you need to archive large amounts of information. The connectors enable you to perform data extraction on different systems and insert them into an archive to store and keep your data safe and secure.

We can also help you manage vast amounts of rich media. OpenText Knowledge Discovery Connectors support human information processing, including audio, video, images, and more. You can optionally and seamlessly perform actions on speech and images to understand rich media in a scalable and sophisticated way via implementation of functions such as speech-to-text, OCR, and face recognition.

Extract and apply native security schemes to document indices. OpenText Knowledge Discovery Connectors maintain and honor native security in real time to facilitate accurate updates or changes to security permissions across individual users and groups.

Key benefits

- **Intelligent data discovery:** AI dynamically guides the discovery of relevant data sources to increase efficiency and effectiveness.
- **Proven consistency:** Automate the processing of vast amounts of data in a reproducible, consistent, and deterministic manner.
- **Unparalleled coverage:** With access to hundreds of repositories, you're provided extensive coverage of connector repositories and their various versions.
- **Straightforward OEM integration:** With easy configuration, flexible API choices (.NET, C, and Java), and a cross-platform architecture, implementing OpenText Knowledge Discovery Servers and Connectors on your own systems couldn't be easier.
- **Rich features for countless applications:** The range of connectors offer the capability to reach content in more repositories, gather the value of that existing content, enrich it and build highly capable applications.
- **Minimal impact:** OpenText Knowledge Discovery Connectors, where possible, use native APIs to interact with third-party repositories, minimizing development time and disruption.
- **Save time, resources, and money:** Writing your own connectors requires the time, human resources, and high costs associated with supporting and maintaining a large library of connectors.

Key features

- **Dynamic Corpus:** Enables unprecedented information access with dynamic data-led discovery of relevant web sources including the Dark Web, which might otherwise be inaccessible.
- **Information governance:** Manipulate content inside the repository by inserting data or disposing of it, all performed via the native API or SDK for easy maintenance.
- **Bi-directional connectors:** Don't just extract data, enrich repositories by inserting new data or metadata derived from data analytics.
- **ETL functionality:** Extract, transform, or load content from various content repositories; embed LUA script to build your own custom logic and integrate with application logic.
- **Advanced actions:** Access a wide array of advanced actions, including synchronize, collect, move, delete, hold, release, and view. Availability of actions may vary by connector.

Resources

[See a complete list of connectors ›](#)

[Learn more about OpenText Knowledge Discovery ›](#)

- **Rich media:** Understand video, audio, and images, and leverage sophisticated functions, including speech-to-text, OCR, Number plate, vehicle recognition, face recognition, and object classification.
- **Entity recognition:** Intelligently classify information by finding patterns in your data to enrich your repositories. For example, you can use entity recognition to validate social security numbers or detect personally identifiable information (PII).
- **Published and uniform API:** Interact with connectors using a simple, common set of action calls. Performing a legal hold in SharePoint, for example, is the same as performing a legal hold in OpenText Documentum Content Management— simply point to a different connector.
- **Native security:** OpenText Knowledge Discovery extracts and maintains native security schemes to document indices and content, preserving data security.
- **Scalability:** Scale up using configuration or architectural capabilities or components.