

# IBM Watson Knowledge Catalog

---

## Highlights

- Establish common business language and vocabulary
  - Centrally manage reference data
  - Customize data classifiers to automatically understand your data
  - Protect against the misuse of data and AI models
  - Share assets based on organizational policies
- 

## Delivers trusted business-ready data to your enterprise

Organizations are racing to unleash the power of their data. Many have been successful in collecting and storing data across data warehouses and data lakes, but they're unable to trust and use the data with confidence for analytics and artificial intelligence (AI) initiatives. Data challenges are systemic, but are also often specific to roles in the organization.

For example, chief data officers (CDOs) manage data that increases in volume, diversity and complexity. They must ensure that data containing sensitive or personal information is protected. CDOs struggle with staying in compliance with changing regulations because they lack effective governance programs and integrated tools. Roles like analytics leaders, business users and data scientists often find it difficult to access data from disparate sources. They don't always know what the data means and whether it can be trusted since its provenance is unknown. As many organizations discover, relying poor quality data can lead to a negative impact on the business.

What organizations need need is an integrated solution that can create and deliver trusted data through self-service access. They need a cohesive solution that can seamlessly provide data integration, data governance, and data quality and data consumption capabilities to deliver high-quality data that users know, trust and understand. An intelligent solution that can automate time-consuming and labor-intensive data processes to make data more simple to discover, curate and access.

## IBM Watson Knowledge Catalog

IBM Watson Knowledge Catalog is powered by IBM Cloud Pak for Data, an intelligent, integrated data governance, quality and catalog solution that delivers trusted, business-ready data to enterprises.

Watson Knowledge Catalog helps organizations build a trusted data foundation. It streamlines the data pipeline by cataloging structured and unstructured data from disparate sources through automatic data discovery, quality analysis and enrichment. It delivers trusted data, whether it's for regulatory compliance or for advanced analytics.

Advanced artificial intelligence, machine learning and deep learning features allow organizations to be agile with their data and keep up with the fast pace of business. With Watson Knowledge Catalog, organizations can:

- Automate time-consuming processes to speed up data curation, profiling and quality analysis
- Suggest assets to the user based on factors such as past usage and content similarity
- Adjust search result ranking based on user search history

Watson Knowledge Catalog serves as a single source of truth for data engineers, data stewards, data scientists and business analysts to gain self-service access to data they can trust and use with confidence.

## Key Features

### **Build your data governance framework**

#### **Establish common business language and vocabulary**

Speak the same language across your business with a common vocabulary by documenting the concepts and dimensions of the business. Better understand data assets in context by associating data assets and attributes to business glossary terms.

#### **Centrally manage reference data**

Instead application storing separate lists of valid values, manage codes and values in a central location where applications can pull consistently. Automatically profile data to understand which reference data sets are used and how accurately they follow those values.

#### **Customize data classifiers to automatically understand the content of your data**

Define and customize data classes so that structured data can be automatically classified to help users better understand their data. Data classifiers can help protect data based on data protection policies.

### **Protect data and AI model misuse and confidently share assets based on organizational policies**

Active policy enforcement and powerful governance tools control and protect access to data, which helps to prevent misuse. Sensitive data can be dynamically masked so that data sets and models can be shared across the enterprise in a security-rich environment.

### **Automatically enforce workflow when building out a governance program**

Create workflows based on company processes and the Responsible, Accountable, Consulted, Informed (RACI) matrix. All changes to a data governance program will automatically trigger approval workflows and notifications to the right stakeholders.

## **Discover and understand your data**

### **Accelerate data curation with machine learning, profiling and term assignments capabilities**

User Quick Scan or Automated Discovery features can discover data assets from a connected source. You can choose to analyze terms like classification, data quality and business term assignments. You can also select the sample size for the analysis.

### **Data quality analysis**

Evaluate the content and structure of your data for consistency and quality. Watson Knowledge Catalog helps you improve the accuracy of your data by making inferences and identifying anomalies.

### **Explore data lineage and relationships**

Create reports that analyze the flow of data from data sources, through jobs and stages, and into databases, data files, business intelligence reports, models and other assets.

## **Consume your data**

### **Self-service access to data**

A powerful and intelligent global search helps users find what they need across different projects, catalogs and governance entities. Machine learning enhances search suggestions and

results rankings, quickly guiding users to the most relevant assets. This feature helps them spend less time looking for data, and more time performing analyses or building AI and data science applications.

### **Interactively discover, cleanse and prepare your data**

With a built-in data refinery, Watson Knowledge Catalog provides a rich set of capabilities that allows you to discover, cleanse and transform your data with data shaping operations. The solution comes with powerful profiling and visualization tools like charts and graphs to help you interact with and understand your data. Data access policies defined in Watson Knowledge Catalog are enforced in the data refinery to ensure that sensitive data that originated from governed catalogs remains protected.

### **Power data science**

Seamless integration with IBM Watson Studio helps data scientists to quickly drive productive use of their data in a suite of powerful tools. Search for data assets in a catalog, quickly evaluate quality, prep the data using data refinery, then add it to a project within Watson Studio to take advantage of a breadth of open source data science tools. Projects allow users to work on small teams using a subset of data that has been grouped and prepared for a specific objective.

### **Accelerate your data**

Connect with third-party external data sources like Pitney Bowes to bring location data or population data into your analysis.

## **Deployment options**

IBM Watson Knowledge Catalog is available in two different deployment options.

- Watson Knowledge Catalog powered by Cloud Pak for Data includes the most comprehensive and integrated capabilities for data governance, quality and catalog. This private cloud deployment option can be installed on any cloud with Redhat OpenShift.
- Watson Knowledge Catalog on IBM Cloud is the public SaaS offering with a complimentary trial tier. There are two additional paid tiers available. This version provides similar data catalog capabilities as the Cloud Pak for Data version. For data governance it provides a simple business glossary and data protection policy enforcement, like dynamic data masking.

## Why IBM?

IBM offers a comprehensive, scalable, solution with data integration, governance to successfully deliver trusted data for insights and compliance to businesses, governments and individuals.

## For more information

To learn more about IBM Watson Knowledge Catalog, visit: [ibm.com/cloud/watson-knowledge-catalog](https://ibm.com/cloud/watson-knowledge-catalog).

Try Watson Knowledge Catalog on IBM Cloud in a complimentary trial:  
[datapatform.cloud.ibm.com/registration/stepone?apps=data\\_catalog&context=wdp](https://datapatform.cloud.ibm.com/registration/stepone?apps=data_catalog&context=wdp)

Learn more about Unified Governance and Integration at [ibm.com/unified-governance-integration](https://ibm.com/unified-governance-integration). Follow us on Twitter at @IBMAalytics, on our blog at [ibmbigdatahub.com](https://ibmbigdatahub.com) and join the conversation #IBMUGI.

Additionally, IBM Global Financing provides numerous payment options to help you acquire the technology you need to grow your business. We provide full lifecycle management of IT products and services, from acquisition to disposition. For more information, visit: [ibm.com/financing](https://ibm.com/financing)

© Copyright IBM Corporation 2019.

IBM, the IBM logo, and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at <https://www.ibm.com/legal/us/en/copytrade.shtml>, and select third party trademarks that might be referenced in this document is available at [https://www.ibm.com/legal/us/en/copytrade.shtml#section\\_4](https://www.ibm.com/legal/us/en/copytrade.shtml#section_4).

This document contains information pertaining to the following IBM products which are trademarks and/or registered trademarks of IBM Corporation:  
IBM®, IBM Cloud™, IBM Cloud Pak™, Watson®

---



All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.