

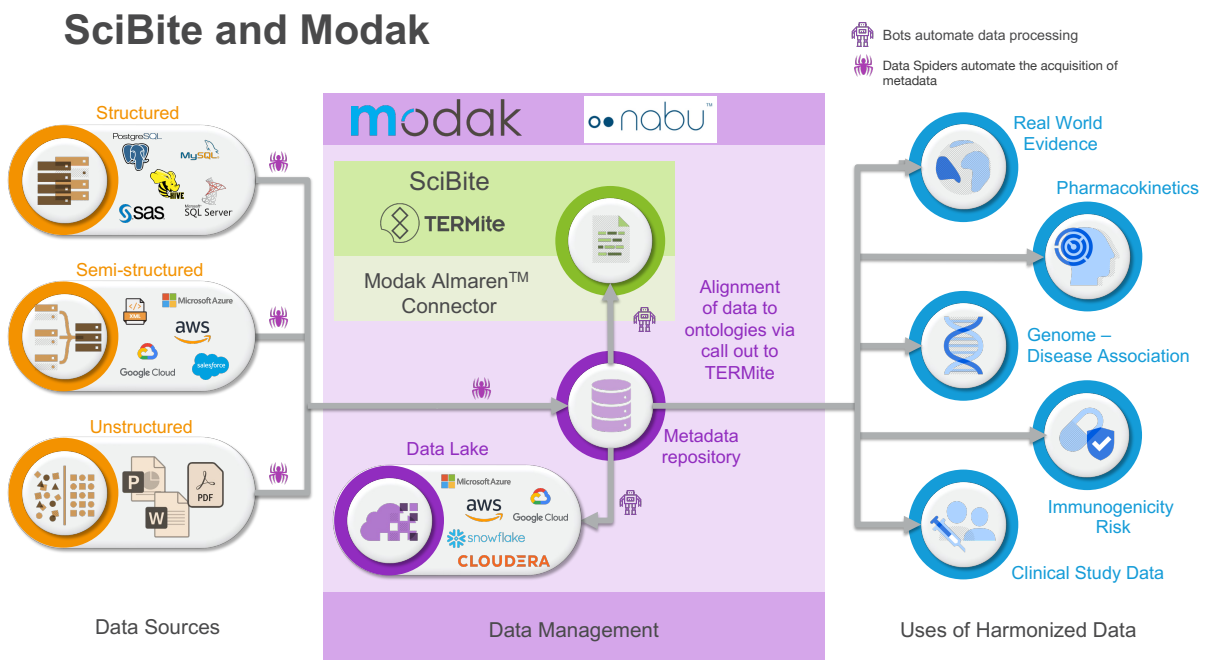
SciBite Partnership with Modak

Partnership Overview

Modak and SciBite are proud to work together with a joint mission to expedite the generation of insights from research publications, patents, and documents; crucial to advancing scientific discovery.

Modak’s data orchestration platform, Modak Nabu™, enables enterprises to automate data ingestion, curation, and consumption processes at a petabyte scale and within a robust data governance framework. As part of the partnership SciBite’s named entity recognition tool, TERMite, is connected into Modak Nabu™. This connection is made possible by leveraging Almarex; Modak Nabu’s rich connector framework built on Apache Spark.

As a result, TERMite can be run automatically within Modak Nabu™ across on-premise, cloud, and external data sources, allowing for machine-readable FAIR data to be fed to downstream applications.



Benefits of the SciBite and Modak Partnership

The integration of Modak Nabu™ with SciBite’s NER capability, TERMite, will empower Life Sciences customers with the ability to:

- Streamline and accelerate the preparation of machine-readable and FAIR data
- Present a more persistent approach to data lineage by maintaining records of data flow between source and target
- Enables TERMite to be called from within a compliant and secure environment for effective data management
- No code approach for creating end-to-end data pipelines leveraging internal and external data sources
- Acceleration in data harmonization and standardization to fuel scientific discoveries

About Modak

Modak is a solutions company that enables enterprises to manage and utilize their data landscape effectively. They provide technology, cloud, and vendor-agnostic software and services to accelerate data migration initiatives. Using machine learning (ML) techniques to transform how structured and unstructured data is prepared, consumed, and shared. **Find out more at modak.com.**

Modak Nabu™ Solution Overview

Modak Nabu™ enables enterprises to automate data ingestion, curation, and consumption processes at a petabyte scale. Modak Nabu™ is a data orchestration platform, combining data discovery, ingestion, preparation, meta-data repository, unification, and profiling. **For more information, visit [Modak Nabu™](#).**

	Active Metadata Catalog	Accelerate data preparation with metadata including technical, operational, social, and business metadata.
	Automated Ingestion Pipelines	Automate building ingestion pipelines in minutes, using active metadata.
	Search and Exploration	Intuitive search and exploration capabilities to discover data, explore relationships, and assess quality.
	Data Standardization	Standardize hundreds of data formats across thousands of datasets in days, using Modak's Data Fingerprinting technology.
	Enterprise Ready	Comply with data privacy, governance, and regulatory requirements with role-based access control, data anonymization, data lineage, & audit trails.

About SciBite

SciBite's data-first, semantic analytics software is for those who want to innovate and get more from their data. SciBite believes data fuels discovery and is leading the way with its pioneering infrastructure that combines the latest in machine learning with an ontology-led approach to unlock the value of scientific content. **Find out more at www.scibite.com.**

SciBite TERMite Solution Overview

TERMite (TERM identification, tagging & extraction) is at the heart of SciBite's semantic analytics software suite. Coupled with SciBite's hand-curated VOCabs, TERMite, can recognise and extract relevant terms found in scientific text. **For more information, visit [SciBite TERMite](#).**

	Rapid Start-Up	Get up-and-running quickly, with no pre-indexing or complex set-up required.
	Robust	Enterprise-grade and scalable to billions of documents, with the ability to run large-scale document processing on systems such as Hadoop.
	Accurate	Precisely tag and disambiguate scientific terms in unstructured scientific text using SciBite's VOCabs containing >20 million synonyms across >80 Life Science topics including genes, drugs, diseases, adverse events.
	Ultra-Fast	Process millions of documents such as the entire Medline database, or large numbers of patent or internal documents in minutes.