

Whitepaper

Data Integration and the R&D Organization

Understanding and Realizing Data's Value in the Enterprise



Data fuels effective R&D. Simply put, data is the lifeline between R&D and breakthrough innovations.

As more data becomes available, it becomes increasingly complex for organizations to enable easy access to information. Despite advances in content aggregation and search, silos limit the free flow of data and analytics. These silos can stem from geographical limits to compatibility issues, to issues of privacy and rights to share information.

In this paper, we're looking at data integration — the process of combining information from different sources — and why it's a valuable approach across the enterprise.

Types of data

When you approach your organization's data, the first challenge will be combining different data formats. Every organization generates and works with a spectrum of structured, semi-structured, and unstructured data.

- **Structured data** is what's stored in rows and columns within a relational database. When a person looks at a spreadsheet that's full of numbers or other data, he or she is typically able to understand the significance of the measurements by reading the data in the chart. Computers, generally, can understand this data, too.
- **Semi-structured data** has an overall organizing scheme that can define a certain level of meaning and provenance. For a machine, understanding semi-structured data may depend on the level to which the data is organized to meet the business need; computers may need instructions in order to properly parse this information.
- Conversely, **unstructured data** has no defined organization, meaning it comes in many varieties and forms (think emails, presentations, documents, video, audio, graphics, and web pages.). For computers, understanding unstructured data, like a text document, is far more difficult than understanding a spreadsheet. Instead of being able to conceptualize what a word means, computers see strings of letters.

Inconsistency between these unlike types of data makes it challenging to unite them in a meaningful way. While a structured data field might appear under a column heading of "Company Name," unstructured data from a series of meeting notes or competitive intel may contain variations on company names that will need to be parsed and assigned to the appropriate "Company Name" column.

What to look for in a tool:

- **Aggregated Search** Go beyond federated search to select and display the data most meaningful and actionable to the user.
- **Contextualization** Enable knowledge workers to personalize their experience and fine-tune the precision of their search results.
- **Open Integration Ecosystem** Bring together publicly available resources, licensed third-party data sources and internally-created proprietary information, leveraging preferred technology and data providers.
- **Semantic Enrichment** Explore information relationships with dynamic visualizations to accelerate insights.

F indable
A ccessible
I nteroperable
R eusable

Data silos: People and processes

Information silos are common within organizations, because each department typically collects its own needed set of data, using purpose-built systems. For instance, marketing teams collect materials like market research reports, competitive intel, and meeting notes, while research strategy teams accumulate journal articles, grant funding info, and clinical trial results, and business development teams monitor the latest developments in patents and start-up companies. All this content is housed in a set of disjointed, incompatible internal and external repositories.

When employees must consult multiple sources to find all the data they need — it results in wasted time and money. In fact, more than 8 in 10 workers worldwide are forced to reacquire or recreate lost documents previously in their company's possession, according to the 2019 Global Intelligent Information Management (IIM) Benchmark report.¹

Integrating data can help free the flow of information and analytics across an organization. But this can be exacerbated by the fact that all this data can reside on a diverse array of private or public cloud-based and on-premise systems. In fact, many organizations even store recent data separately from historical data, further complicating integration efforts. Sometimes data silos are years in the making, so it takes time and a solid strategy to deconstruct them.

FAIR-ifying data

Integration is the first step to making scientific data findable, accessible, interoperable, and reusable — commonly referred to as the “FAIR” guiding principles for researchers and publishers.²

Application of the “FAIR” principles helps enable organizations to maximize the value of their digital assets. After all, knowledge managers can't maximize the value of their digital information investments when knowledge workers can't easily find needed data. Data integration brings awareness of available information resources, promoting discoverability and increasing usage. From a licensing and information management perspective, this optimizes the usage of licensed content and data. From a researcher perspective, data integration is the first step to address challenges of the research workflow, such as the risk that users are ‘missing out’ on vital information or connections. These day-to-day research inefficiencies result in a less efficient research and development pipeline.

Creating a knowledge supply chain

Data integration, and the “FAIR-ification” of it, is a critical step in supporting what we refer to as the “knowledge supply chain.”

The everyday challenges this white paper has highlighted — to access, process, and store data — are not the final objective. After all, the market value of all this information is directly related to the extent that it can be used to alter business outcomes.



Which begs the question: how can an organization make the best use of available data and thereby maximize the utilization of knowledge for its benefit?

To learn more about this concept, we invite you to download a complimentary copy of [The Concept and Importance of Knowledge Supply Chains](#).

How CCC integrates data

The ideal data integration approach maximizes an organization's digital asset investments by unifying multiple data sources within an open integration framework, enabling knowledge workers to find relevant information. To that end, CCC takes a flexible approach to integrating data, depending on the licensing scenario and technology available.

CCC works with data sources to provide semantically enriched content across the application. This includes working with internal proprietary content and external publicly-available and commercially-licensed content sources.

In other words, if you have information in a database or any type of structured digital data, chances are we can provide an integration path, and ultimately make it more easily discoverable.

References

- ¹ <https://fts-eu.com/resources/the-2019-intelligent-information-management-benchmark-report/>
- ² <https://www.go-fair.org/fair-principles/>



Copyright Clearance Center (CCC)

A pioneer in voluntary collective licensing, Copyright Clearance Center (CCC) helps organizations integrate, access, and share information through licensing, content, software, and professional services. With expertise in copyright and information management, CCC and its subsidiary RightsDirect collaborate with stakeholders to design and deliver innovative information solutions that power decision-making by helping people integrate and navigate data sources and content assets.

Learn more about our licensing, content, and data solutions:

U.S. organizations:

🌐 copyright.com/rightfind-navigate

✉ solutions@copyright.com

Outside U.S. organizations:

🌐 rightsdirect.com/rightfind-navigate

✉ solutions@rightsdirect.com