



## Unlocking the Value of **Unstructured Data**

In today's hyper-connected, digital-first world, almost everything we do is tracked, collected and packaged as data. From the vegetables we buy at the grocery store to the temperature we set our smart thermostat in the evening, with the rise of big data, there's no such thing as extraneous information.

As the amount of data in the world continues to multiply at breathtaking speeds, [companies in virtually every industry across the globe](#) are increasingly relying on that data to inform major strategy and operations decisions.

While most organizations focus their data analysis efforts on structured data—data that follows conventional data models and can be stored and managed in a traditional database—[the vast majority of data available today is actually unstructured](#). Despite ongoing innovations in artificial intelligence, machine learning and other technology, unstructured data has remained largely overlooked and misunderstood in the data science community. Although its cumbersome nature has given it a bad rap, with a little work unstructured data can be an incredibly powerful resource—and provide a distinct competitive advantage for companies that understand how to use it.

**In this eBook, we'll go deeper into the differences between structured and unstructured data, how your organization can unlock the value of unstructured data, and some examples of unstructured data's real-world impact.**

*With the rise of big data, there's no such thing as extraneous information."*





The growth of global data is estimated to skyrocket to 180 zettabytes (180 billion terabytes) by 2025, with 80 percent of it unstructured.”

- [Automation Hero](#)



# Defining the Three Categories of Data

## 1. Structured Data

Traditional **structured data**, like financial transaction data and retail inventory, adheres to a rigid format to ensure consistency when it's being processed or analyzed—think: columns and tables, spreadsheets and databases. It has clearly defined internal parameters and primarily contains numerical information. Since it's organized in this format, structured data is considered much easier to search, refine and analyze by both humans and machines, often by using structured query language (SQL) to navigate the datasets more effectively.

1

STRUCTURED

2

SEMI-STRUCTURED

3

UNSTRUCTURED

### EXAMPLES OF STRUCTURED DATA

- Customer Contact Information
- Financial Transactions
- Retail Inventory
- Airline Reservation Systems
- Database Records
- Online Forms



# Defining the Three Categories of Data

## 2. Semi-Structured Data

Falling in the middle of the road is **semi-structured data**. Much of the information that's labeled as unstructured data could actually be considered semi-structured data, making the term "unstructured data" a misnomer in some contexts. Although unstructured data doesn't follow a traditional data model or schema, even the most disorganized information generally still has some form of internal structure.

*"Much of the information that's labeled as unstructured data could actually be considered semi-structured data."*

For example, while the text of emails would be considered unstructured, each email typically includes internal structure from predefined metadata fields like who the email was sent to, who it was sent from, time it was sent, subject line, and so on. Other common types of unstructured data like social media posts and text messages follow the same pattern. So, while the content of a Facebook post would be considered unstructured data, the date and time it was posted, and the number of "likes" or "shares" it got would be considered semi-structured data.

Because of this, unstructured data and semi-structured data often work together, and the terms are sometimes used interchangeably for simplicity's sake.



### EXAMPLES OF SEMI-STRUCTURED DATA

- Social media "likes"
- Number of video views
- Email metadata
- Word doc metadata



# Defining the Three Categories of Data

## 3. Unstructured Data

Unlike structured data, **unstructured data** is information that does not follow an organized format or predictable order. It can be human generated—like text messages, emails, social posts, and so on; or machine-generated—like information collected from IoT devices, for example. It's usually text-based and qualitative instead of quantitative.

*“Unstructured data is usually text-based and qualitative instead of quantitative.”*

The vast majority of new data being created on a daily basis is unstructured. With the spread of the global pandemic and more people than ever working, playing and ordering their basic necessities from home, the volume of unstructured data has exploded in the last couple years. The messaging platforms (like Slack or Teams) you use to communicate with your co-workers during the day? That's unstructured data. The videos you stream after work to relax while you eat your dinner? Also unstructured data. The websites you click around on when trying to decide which new athleisure suit you want to order? Yep, you guessed it: unstructured data. It may not follow the easily recognizable formats and schemas of traditional data, but it's still incredibly valuable information—and it's available at your organization's fingertips.



### EXAMPLES OF UNSTRUCTURED DATA

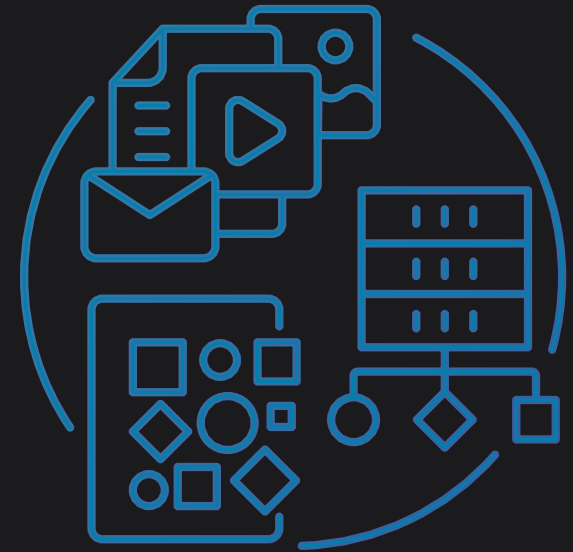
- Text files and documents
- Audio files
- Video files
- Images
- Email text
- Social media posts/messages
- Chat logs
- Website content



# Four Defining Features of Unstructured Data

In addition to new unstructured data being created every moment, there's also an endless array of data already out there waiting to be utilized, ranging from historical news coverage to past company communications.

**Although it comes in many different forms, there are four defining features of most unstructured data:**



## 1. It's textual

The majority of unstructured data consists of textual information, like news stories, company communications, legal communications, and more.

## 2. It's unwieldy

Unstructured data is essentially large amounts of information that have been aggregated or dropped into a system without any sort of obvious formatting or organization, making it more difficult for analysts to put to use.

## 3. It's qualitative

Unlike the hard numbers and financial information that data analysts are used to working with, most unstructured data is qualitative in nature. This can make it tricky to interpret, but also offers valuable opportunities for exploring new kinds of analysis.

## 4. It's contextual

Unstructured data like news coverage and biographical information is extremely helpful in providing analysts with the context to go beyond basic facts and numbers to understand why something is happening.

# Smart Data Enrichments from Nexis Solutions

## Article and Entity Sentiment

Tags for overall article sentiment and sentiment detected on mentioned entities or a person

**EXAMPLE:** Search for articles with negative sentiment on a company

## Entity Extraction

Named entity extraction of companies, organizations, places, and people

**EXAMPLE:** Search for articles about an organization or location

## Key Article Metadata

Article date, title, author, length, and load date

**EXAMPLE:** Search only articles with more than 500 characters

## Key Financial Identifiers

Industry codes (NAICS and SIC), as well as information about public companies mentioned, including the stock symbol and exchange in which it trades

**EXAMPLE:** Search for articles on an industry using standard industry codes

## Key Source Metadata

Publication name, date, and media type

**EXAMPLE:** Search by publication, date, or media type

## Media Value

Approximation of the monetary value of the entire story

**EXAMPLE:** Search for articles citing your brand to understand the potential monetary value of press coverage

## Print Circulation

Industry standard for measurement of print reach

**EXAMPLE:** Search only in print publications with more than 500K circulation

## SmartIndexing

Subjects and industry tags

**EXAMPLE:** Search for articles on a specific industry or filter for business news by excluding celebrities or sports

## SmartIndexing

Places, people, and companies

**EXAMPLE:** Search for articles about a newsworthy individual

## Source Category

Type of publication an article is sourced from

**EXAMPLE:** Search by type of publication (news, companies and financials, cases, biographies) and sub-category (blog, company profile, M&A, newswire, etc.)

## Source Location

Country (or county, city) location of the publisher of the source

**EXAMPLE:** Search by publisher location

## Source Rank

Editorial ranks are applied to news sources only

**EXAMPLE:** Search only top international, national, and business news sources or exclude consumer sources like The Onion

## Source Section

Information about the section in source from which the article derives

**EXAMPLE:** Exclude "Obituary" or "Opinion"



# How to Unlock the Value of Unstructured Data

In its raw form, unstructured data isn't particularly useful to data analysts or C-suite executives. But it can play an important role in a company's data analysis process and strategy—and lead to real-world impact—if leveraged correctly.

**There are three key steps required to unlock the value of unstructured data:**

## STEP 1:

### Curation

Curate the data to only include vetted and trusted sources that won't lead to false conclusions. In a world inundated with made-up news and information [where only 53% of people trust the media](#), this step is more important than ever.

*"In Q3 of 2020, there were **1.8 billion** fake news engagements on Facebook."*

*Study conducted by non-partisan United States public policy think-tank, The German Marshall Fund*

## STEP 2:

### Context

Add context that makes the data more valuable by applying specific, detailed enrichments that convert unstructured data to semi-structured, searchable data. Enrichments might include things like geographical coordinates, article subject/topic, article sentiment, languages and so on. Metadata can then be applied to clean, semi-structured data, enabling users to quickly filter out irrelevant data and get the information they need.

## STEP 3:

### Quality Analysis

Conduct quality, in-depth analysis that delivers powerful, actionable insights. This is the most crucial step in the process, potentially leading to measurable results for your organization. For example, say a company experiences a significant dip in their stock price. By analyzing the volume and sentiment of news coverage during the time period surrounding the dip, analysts can better understand why the dip may have happened. Executives and investors can then use that information to fuel smarter, more strategic decisions that may prevent similar dips in the future.



Predictive analytics, machine learning algorithms and other artificial intelligence applications won't run on empty. [Nexis Data as a Service](#) offers bulk, constant call data APIs and on-premises applications for delivering highly relevant, archival and current datasets to power your big data projects. Our expansive source universe features petabytes of semi-structured, normalized and enriched text-based data to meet a wide range of data needs.

- **36,000+** licensed & **100,000+** total news sources in **37** languages from **200+** countries
- News archive spanning **45+** years
- **2.7M+** web / social media sources in **37** languages from **200+** countries
- **1.5M** profiles of PEPs, their family members, and close associates
- **1,400+** sources of domestic and global sanctions, watchlists and blacklists
- **400M+** Company Intelligence records
- Legal entity data covering **165M** companies and **220M** officerships in **130** jurisdictions
- **100+** patent authorities

# Unstructured Data's Real-world Impact

“In the mid-2000s, retailers were the first to combine and analyze data from customer emails, images, voice, and store-traffic records to market to particular customers. And in the last few years, finance has made substantial progress,” writes Tam Halbert of the Massachusetts Institute of Technology (MIT). “Now other industries, including shipping, transportation, legal, and real estate, are leaning into unstructured data.”

At LexisNexis, we work with hundreds of companies that leverage unstructured data to inform their business strategy and decision-making, from educational institutions and non-profits to multi-national corporations.

**Let's look at three real-world examples of unstructured data in action.**



## 1. Financial Institution

### DATA

News articles, PEPs, Sanctions, company reports

### SAMPLE ENRICHMENTS

Article subject, company data (e.g., revenue, stock ticker), article relevancy

### OUTPUT

Risk assessment, adverse news monitoring, portfolio management

All of the financial institutions LexisNexis works with are required to carry out due diligence and risk assessments before they do business with other companies. Increasingly, these financial institutions are embedding our data and enrichments into their due diligence workflow and process automation to help assess risks and comply with financial regulations.



## 2. Oil and Gas Company

### DATA

Traditional and social media, earnings calls, new patent announcements, product-related announcements

### SAMPLE ENRICHMENTS

Article subject, company data (e.g., revenue, stock ticker), entities mentioned

### OUTPUT

Competitive & market analysis, trend forecasting

An oil and gas company we work with uses a number of different LexisNexis content sources and enrichments to monitor news in their marketplace. Analyzing this unstructured third-party data—like news, announcements and social media—allows them to better understand how trends evolve, how they are covered in the media and how they might be predicted in the future.



## 3. Luxury Fashion Brand

### DATA

News articles, press releases, social media

### SAMPLE ENRICHMENTS

Article subject, entities / brands, reach, sentiment

### OUTPUT

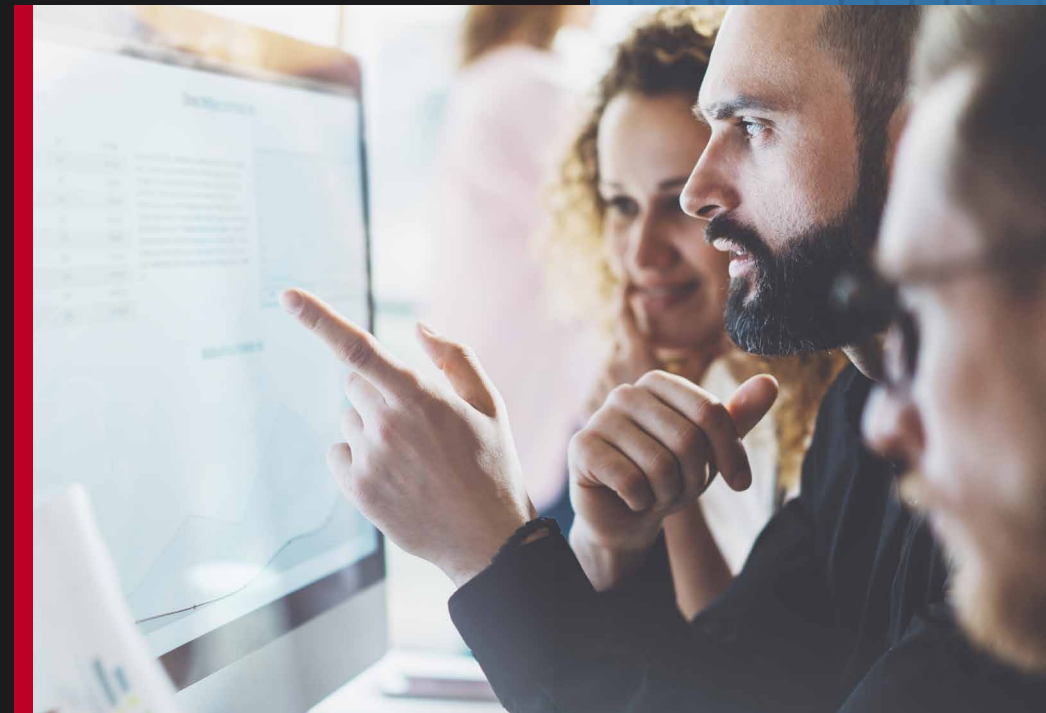
Media monitoring, public relations & brand insights

One of our clients, a luxury fashion brand, needs to stay on top of traditional media coverage and, more importantly, social media coverage and sentiment. By leveraging data enrichments like mentions, product placement, sentiment and reach, they can get a more accurate understanding of their brand's reputation, as well as endorsements and coverage by high profile social influencers.

# How to Get Started with Unstructured Data

If your organization isn't tapping into the value of unstructured data, you might be missing out on key insights and information that could help inform critical business decisions in the future. Not sure how to get started? [Nexis Data as a Service](#) can help.

Data lies at the heart of digital transformation. Whether forecasting with predictive analytics or conducting trend analysis, improving competitive intelligence or ensuring accuracy and compliance, we provide the complementary third-party data your organization needs to make strategic, data-driven decisions. Our expansive source universe features petabytes of semi-structured, normalized and enriched text-based data to meet a wide range of data needs.



## About Nexis® Solutions

Nexis Solutions, a division of LexisNexis, is dedicated to developing innovative tools to support data-driven decision-making. Our commitment extends beyond comprehensive content and outstanding search technology to world-class client service support, ensuring that our clients gain maximum insights and value from LexisNexis solutions.

## Arrange a demo to learn more

[LexisNexis.com/DaaS](https://www.lexisnexis.com/DaaS) | [1-888-466-3947](tel:1-888-466-3947)

