

CITO Research

Advancing the craft of technology leadership

The Automated Analyst: Transforming Data into Stories with Advanced Natural Language Generation

SPONSORED BY

NarrativeScience 





CONTENTS

<u>Introduction</u>	1
<u>The Gap Between Data and Decisions</u>	2
<u>A New Approach: Using Advanced NLG to Help Data Tell Its Story</u>	3
<u>Natural Language Generation: It's Not All the Same</u>	5
<u>Narrative Science and Quill™</u>	8
<u>Narrative Science in Action</u>	8
<u>Conclusion</u>	10



Introduction

Executives don't usually do their own analysis. They have expert staff that provide them with briefings and updates about particular areas, feeding them critical information and concise summaries that tell them what they need to know and its impact. The executives have expertise in strategy and decision making and depend on staff with expertise in data and analytics.

This CITO Research white paper posits that because of the proliferation of data and its relevance to everyone's jobs, we all need these types of briefings. We don't need more data, more spreadsheets, or even more beautiful visualizations. What we need are the key takeaways: a way to understand the impact of the story the data is telling now. We need an analyst at our elbow, ready to provide us with that level of information in a quick consumable form. We need information that is tailored to our domain and to our particular roles. We all need this type of information when we want it and need it, a reality that is only achievable with machine scale.

Advanced natural language generation closes the gap between data analysis and informed decision-making

Advanced natural language generation is an important technology that turns copious data into a summary that enables you to read, digest, and understand what is happening and its impact on you. And when you have questions about what it says, you can trace every detail right back to the system of record. Importantly, this technology does not just create data stories for executives, but for each regional sales manager, each store manager, the finance department, the warehouse staff, the purchasing department, and for as many audiences as you would like to arm with key information. And unlike people, this technology does scale.

Advanced natural language generation (Advanced NLG) closes the gap between data analysis and informed decision-making. This CITO Research paper examines how NLG has evolved and how Advanced NLG can help companies get consumable, actionable stories from their data.



The Gap Between Data and Decisions

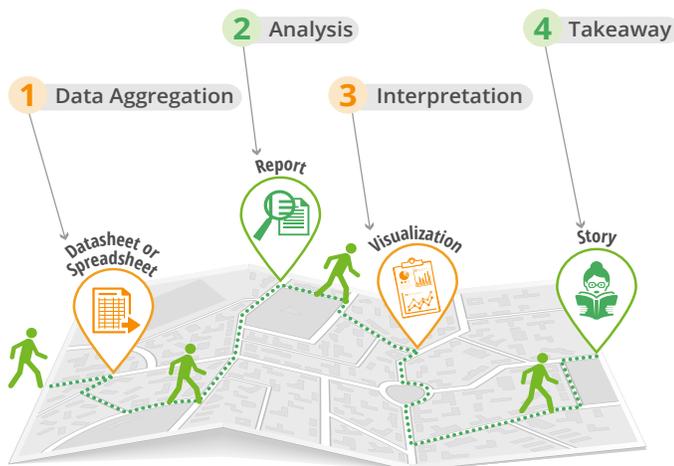
At best, data and analytics technologies aggregate, analyze, and visualize data. Even so, an important gap remains: the story of what that data means. The problem is that companies don't have a way to efficiently bridge this gap and convey the most important and interesting information within the data to benefit people all across the business.

Typically, companies take a manual approach to interpreting data and sharing insights. Data scientists and analysts are hired to use analytics tools to create models, reports, graphs, and charts. But even data visualizations do not go far enough to communicate what the data is really saying; visualizations still require interpretation. Business stakeholders often need analysts to provide them with key insights and provide a written analysis (or PowerPoint) of their findings.

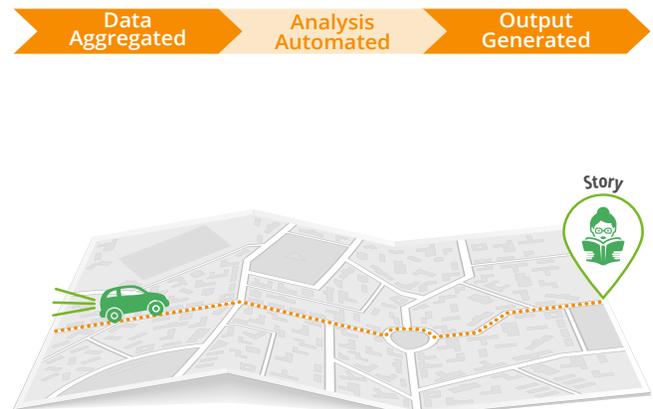
This approach is problematic for several reasons. For starters, it is incredibly time consuming. The business loses a potential first mover advantage every time data must be interpreted. Additional data scientists and analytics staff help to some extent, but hiring personnel skilled in data literacy is difficult. Such talent is in short supply (and therefore expensive) and finding people with deep analytics skills who also know your business well can be very challenging.

Because of these disadvantages, as well as the difficulty of engaging customers and business users with relevant, personalized content, data insights are underutilized. Without time or resources for interpretation, too often data can't be effectively factored into important decisions. As a result, companies lose business opportunities and do not realize a return on their considerable analytics investments.

There is a better way: closing the last mile of data analysis with Advanced NLG.



Manual, human-intensive, inefficient data analysis and interpretation



Automated, machine-driven, efficient data analysis and communication



A New Approach: Using Advanced NLG to Help Data Tell Its Story

Companies can operationalize data insights within their organizations by leveraging the power of Advanced NLG fueled by artificial intelligence. Advanced NLG platforms automatically transform numbers (structured data) into readable summaries with key takeaways in plain English, thereby addressing the challenges associated with manually interpreting data. It also enables the creation of new knowledge products by transforming data into high quality, relevant communications at scale. Let's look at each of these characteristics.



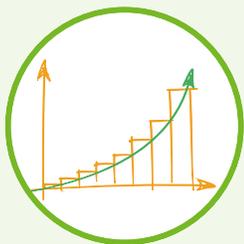
Speed

Because they are automated, Advanced NLG solutions rapidly deliver the information business users and consumers need, giving companies a first mover advantage. Intended audiences—whether employees, partners or customers—receive accurate, insightful summaries in seconds, at a speed only possible when delivered by a machine.



Quality

Advanced NLG systems aim to take the work of your best analysts and brand ambassadors and consistently deliver the same high-quality content, on message, every time. The resulting data-driven narratives effectively replicate the data insights of your best analysts as well as the voice, style, and word choice of your best brand representatives. Automation eliminates errors and delivers high quality insights and informative summaries that business users rely on to drive decision-making.



Scale

Advanced NLG codifies the way that experts discover, interpret, and communicate insights, enabling everyone, not just executives, to get briefings of what they need to know. Instead of needing to train new employees on processes for measuring the health of the business, Advanced NLG provides key performance indicators and drivers in plain language. Analysts and data scientists are no longer bottlenecks in the process but instead can apply their advanced skills to exploratory work to identify models and determine what a domain means, making their work more interesting and valuable.



Relevance

A key part of analytics is not just knowing what the data says or even what the data means, but what it means to the person who wants the story. Advanced NLG tailors data for particular audiences, ensuring that the information is relevant to the person's role. Rather than displaying numbers and charts, Advanced NLG deepens customer and employee engagement by delivering insights in plain language.

Advanced NLG solutions ensure relevance in three ways:

- Identify information that is important to a particular business domain.
- Refine domain information to meet the needs of stakeholders. (What is important to the finance department is different from what's important to customer service.)
- Customize narratives targeted for a specific person's needs. (Some users want a short summary, while others want a long report with more details.)

In practice, that might be the difference between a mobile alert for an investor who wants to track his performance or a 15-page in-depth report for an investor who prefers a more detailed analysis.



Traceability

When anyone performs analytics, people have questions. Where did the numbers come from? How did you arrive at that conclusion? Why do you think we should take that course of action based on your analysis?

When computers perform advanced analysis, people have even more questions. And some systems that perform such analytics are a black box of sorts, lacking transparency into the analytical process and the results.

Advanced NLG offers traceability that provides a level of trustworthiness. Traceability enables users to understand why the system chose to communicate data insights in a particular manner. The system provides evidence not only for its overall advice and framing of the story, but can also explain why it chose specific words to tell the business story.

Most importantly, for questions, for auditing, and for compliance, each word and insight can be traced directly to data and to systems of record, offering complete visibility and traceability. It is for this reason that Advanced NLG is already being used by highly regulated industries such as financial services as well as by government entities, including the intelligence community.



Advanced NLG and Innovation: Increasing Customer Engagement

Innovative companies are realizing that they are sitting on untapped or underutilized data that offers tremendous value to others. They are packaging up that data and presenting it to customers in a way that is easily consumable.

This is where Advanced NLG solutions from Narrative Science come in, helping companies deliver high-quality information products at scale.

Dominion Dealer Solutions uses Narrative Science Quill to create a product called **StoryBuilder** that draws data from numerous sources to create a unique story for each vehicle shown online. StoryBuilder draws on all of the basic information about a vehicle (its age, number of owners, features, make and model) and enriches that with numerous data sources, including automotive reviews, CARFAX history, and pricing relationship to books such as Kelley Blue Book. It then generates dynamic descriptions highlighting key features that viewers care about, increasing engagement and time to market.

In fact, according to an independent study of 304,000 vehicles with StoryBuilder descriptions versus more than 3.1 million vehicles without such a description, cars with a StoryBuilder description sell faster: new vehicles sell 10.3% faster, certified vehicles sell 22.3% faster, and used vehicles sell 34% faster, in an average of 39 versus 59 days. With effective data-driven stories for the cars they have in stock, dealers can sell more cars, faster.

Credit Suisse utilizes Quill to enhance one of its investment research products, HOLT. The HOLT platform is an objective framework that compares and values 20,000 companies from around the world. Quill incorporates hundreds of variables from these companies, including asset growth, risk and performance momentum and generates narratives that clearly explain a company's performance, offering a tangible explanation to accompany the platform's existing charts. The narratives are updated in real-time to make sure they always reflect the latest in a company's performance. By utilizing Quill, Credit Suisse HOLT has improved investment research coverage, quality and consistency.

Natural Language Generation: It's Not All the Same

Natural language generation is not new, but it has come a long way.

The simplest level of NLG takes a few data points and turns them into sentences. A simple weather report example is a sentence like this: "the high today will be 72 degrees."

The next level of NLG takes a templated paragraph and generates language based on the changing data. Sports scores often can be handled this way. For certain types of templated reports, this works. The analytics performed by such systems isn't very sophisticated (it's generally driven by business rules framed as if/then statements).



Advanced NLG transforms data into a narrative with a beginning, middle, and end. This narrative, or story, is based on an in-depth analysis of the data. For example, suppose you want to produce a quarterly earnings analysis to be shared with your board of directors. Transforming the data into a narrative will take into account four levels of information:

1 Purpose

What you want to say.

Are the company's earnings growing or on the decline?



2 Information

What you need to know.

This quarter's earnings compared with last quarter's.



3 Analytics

How you figure it out.

What type of analytic is best suited to the problem? In this example, it's time series analysis; in another case, it might be regression analysis or predictive analytics.



4 Data

What data you need.

Quarterly earnings and projected earnings (along with the level of confidence in those projections and possibly some alternative models that consider different variables that could impact earnings).



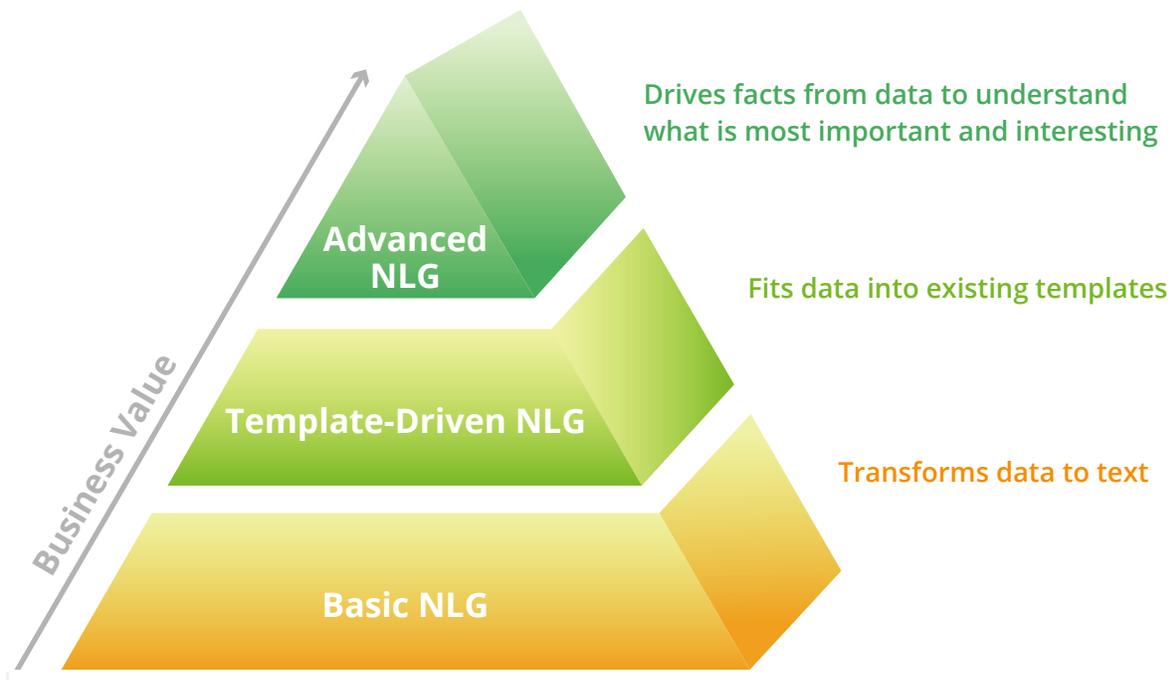


The narrative produced is tailored to the needs of, in this case, the board. Most of the board members want a 5-page summary, but one likes more detail, with earnings broken down by product line, division, and sales channel.

In order to create a narrative from the data, all of these needs must be taken into account. The communication goal defines what analysis must be performed in order to get to the truth of the story. The analysis both serves the story and is defined by the story. The result is an easily consumable narrative that puts data insights in context for a particular audience with the added value of actionable recommendations for improving the scenario.

The result is an easily consumable narrative that puts data insights in context for a particular audience

Advanced NLG uses artificial intelligence and a domain model to determine the most interesting and applicable data insights. If the system is analyzing ten factors that affect a supply chain, and eight are relatively stable while two show rapid change, the system will report on the two showing change. In addition, one domain model can create multiple renderings so that different stakeholders get a version of the story that is applicable to their specific needs.





Narrative Science and Quill

Narrative Science delivers data-driven communications at machine scale with its Advanced NLG platform, Quill. Quill bridges the gap between data and the people who need to understand it. It starts by understanding the user's communication goals and the data and appropriate analytic techniques to meet those goals. It then analyzes the context of what is being said, and decides what pieces of information are most interesting and important to communicate. The result is a customized natural language narrative. All the recipient has to do is read.

Quill bridges the gap between data and the people who need to understand it

Narrative Science in Action

Writing Customer-facing Investment Fund Reports

An investment company provides a quarterly performance commentary to clients on each of its 200 core mutual funds and 100 separately managed accounts. This previously required the efforts of more than ten writer-analysts working full time for about a month. With the goal of increasing writer-analyst productivity, the company decided to implement Quill to automate the language and analysis for all funds in the portfolio—every quarter, every month, and on demand. Quill's commentary integrates directly into the workflow, giving writer-analysts immediate and on-demand access to commentary. In addition, Quill allows the company to focus more on thought leadership content and has shortened time to market by up to 30 days.

Automating Internal Sales Reports

One of the nation's largest insurance firms was manually creating scheduled and ad hoc reports for regional sales managers. Each report took at least two hours to create, which hampered the firm's ability to scale reporting and provide relevant, timely updates to different stakeholders without adding staff. The firm sought a way to reduce the time and effort spent digging through sales data and writing reports, and it wanted to quickly and consistently update internal stakeholders on key performance metrics with new sales opportunities with plain language summaries. To this end, the firm implemented Quill.

Quill writes hundreds of on-demand regional reports for the management team and uses the same data set to generate detailed scorecards for 4,000 individual field sales representatives.



Automating the Generation of Suspicious Activity Report (SAR) Narratives

On average, financial institutions receive over 1 million alerts annually that they are legally required to investigate within 30 days and, if warranted, file Suspicious Activity Reports (SARs). The anti-money laundering (AML) team of the financial services institution sought a way to generate the narrative component for every SAR to meet the necessary written regulatory requirements and highlight trends and patterns in the types of suspicious activity.

Quill generated SAR narratives from AML data detailing the “who, what, when, where and why” of the situation, streamlining the investigation process and greatly reducing the time spent on investigations.

Experience Quill in Action

Narrative Science has numerous case studies on its website. If you have experience with Google Analytics or an active Twitter feed, you can try out two free Quill applications: Quill Engage and Quill Connect.

Quill Engage takes Google Analytics data and turns it into a report with highlights of what is happening on your website, showing you which pages are drawing the most views and where referrals are coming from (among many other highlights).

If you're looking to better understand your engagement on Twitter, Quill Connect analyzes your Twitter feed and draws comparisons between you and your followers, including a distribution of what topics you (and your followers) are tweeting about this week.



Conclusion

Data literacy, the ability to draw conclusions from our own analysis of data, is an increasingly important business skill. Everyone needs to improve their data literacy, their ability to draw conclusions from observing trends in data and asking questions that can be answered by analyzing data.

Data literacy is a worthy goal, but the truth is that we can't wait. We all need to know what to do with data now. CITO Research concurs with Narrative Science co-founder and Chief Scientist Kris Hammond that someday we will look back at spreadsheets as quaint tools for understanding data "in the old days." We need technology that can scale our use of data, to tell us the data story while at the same time providing us with complete visibility into where that analysis came from so we can trace back each word and each detail to systems of record.

Data monetization—the ability to create information products and new revenue streams from the treasure trove of data

that each organization has—is a vital use for Quill. Do you have information that would help your partners or customers if they knew how they were doing compared with your other partners or other customers? Such type of benchmarking information is often so valuable that given a taste of it, people are willing to purchase a subscription. Taking multiple data sources and combining them to create reports is a perfect use for Narrative Science Quill.

In order to reap the benefits of data, everyone, not just analysts and data scientists, should be able to understand and act on the stories hidden in the data. Companies need to close this gap, and CITO Research believes that Narrative Science's approach to Advanced NLG can effectively do so. Narrative Science enables better, faster decision-making by quickly conveying the most important and interesting information to any intended audience through easy-to-understand natural language narratives.

Contact Narrative Science to learn how Quill can help your company put data insights to work faster for everyone ▶

This paper was created by CITO Research and sponsored by Narrative Science

CITO Research

CITO Research is a source of news, analysis, research and knowledge for CIOs, CTOs and other IT and business professionals. CITO Research engages in a dialogue with its audience to capture technology trends that are harvested, analyzed and communicated in a sophisticated way to help practitioners solve difficult business problems.

Visit us at <http://www.citoresearch.com>