Too many formats, too little time

Financial statements are at the heart of so many different business processes, from lending to audits, investments, and mergers and acquisitions. For example, prospective buyers and borrowers must regularly submit financial statements to financial services firms. Security analysts and investment bankers also rely on financial statements for developing valuation models.

Processing and analyzing financial statements is a big part of these activities. The American Institute of CPAs (AICPA), the Financial Accounting Standards Board (FASB), and other organizations apply rigorous standards for how to handle these transactions, and their efforts have succeeded—for the most part.
What’s lacking are regulations governing transactions, nomenclature, and the preparation of financial statements. Currently, each company is free to create its own discrete items in a chart of accounts, resulting in a high number of nonstandard account labels. Also, accounting standards vary from company to company and country to country, making cross-country evaluation of financial data challenging. To complicate matters further, the financial statements that companies must submit take many different forms, including electronic files, spreadsheets, PDFs, and faxed documents. That makes gathering, extracting, and normalizing the data for credit and risk decisions complex and cumbersome.

Many companies still handle the financial spreading process manually, which is slow and often impacts quality. Not surprisingly, most prefer to turn the whole exercise over to business process outsourcing (BPO) firms. Yet, this complicated and knowledge-based process is ripe for innovation to create a better way to handle financial spreading.

That’s why we created a contextual, traceable product that applies artificial intelligence (AI) and machine learning to streamline and automate this complex process. We’ve dramatically improved quality, reduced costs, and shortened cycle times at global financial institutions where this product is in large-scale use.

Making sense of dark data

Figure 1 illustrates how most firms now gather, extract, and normalize unstructured data. Analysts receive the data in a variety of forms and from a variety of sources. They re-enter the data manually into a commercial or internally created spreading solution. Investment bankers often use Microsoft Excel, whereas commercial lenders tend to use commercial spreading solutions, and many U.S. financial services firms have developed their own spreading systems.

The story so far

The manual spreading process requires the analyst to manually map the reported items to the analyst’s normalization template. Often, analysts must combine several items in the client’s financial statements to fit their financial institution’s chart of accounts. Sometimes, they use the financial statement footnotes about specific line items, such as fixed asset breakups or allowance for doubtful accounts to provide additional insights into the financials of the company.

Each financial statement is unique to the source, company, industry, and country it comes from. Consider these discrepancies:

Figure 1: Current Process
• **Statement form and format variability.** Unique formats that can change without notice.

• **Variation in content.** Non-standard content and taxonomies can force analysts to “club” data into the next-most-appropriate field, without completely understanding the impact. Additionally, revenue and income labels can vary by industry. The same accounting items can also appear in different sections of different statements. Utilities, for example, might appear under either Cost of Goods or Expenses.

• **Country- and industry-specific variations.** What Canada calls Interest Expenses, for example, U.S. GAAP refers to as Other Expenses.

• **Footnotes.** There is no standard for identifying and applying income statement footnotes that contain critical material, such as operating lease information. Similarly, there is no system in place for handling balance sheet footnotes about property, plants, and equipment breakups. There is also no standard for footnotes that capture off-balance-sheet line items, such as the maximum limit available under revolving credit.
Variation in content

Same accounting items appear in different sections in different statements.

Bank charges mapped to “Interest expenses” in Canada
Many factors are driving the push toward automating the spreading process. Scalable technologies such as computational linguistics and machine learning can improve quality, remove inconsistencies, reduce costs, and make it easier and faster to create an audit trail back to source documents for after-the-fact queries. What's more, as the financial crises of 2008 revealed, it can take a long time to process and analyze financial data, which is why Dodd-Frank regulated the amount of time analyses can take.

Cora LiveSpread

An effective AI solution has to process all types of documents, formats, and semantics. Genpact’s Cora LiveSpread™ completely automates the gathering, extracting, and normalizing of data from financial statements, bank statements, income tax returns, and more. This continuously learning product is form, format, and content agnostic. It also allows for complete audits and intermediate results. Analysts can double-click on an extracted or normalized item and trace it all the way back to the document where Cora LiveSpread sourced it.

Automated Receiver

The Automated Receiver accepts financial statements from different sources and in different formats, including PDFs, fax, paper spreadsheets, and data files. Its intuitive user interface makes it easy for clients to upload their statements. It can also automatically extract financial statement data and commentary from accounting systems such as Intuit QuickBooks and Microsoft Dynamics GP, or in enterprise systems such as SAP and Oracle financials.

Automated Extractor

The Automated Extractor mines financial data from source documents, including PDFs. As for image files, Cora LiveSpread first improves their quality using proprietary image processing algorithms. Next, it applies optical character recognition (OCR) to turn the images into machine-readable documents. Then, Cora LiveSpread identifies and extracts the data with domain-independent algorithms and geometric scaling. A metadata structure stores a model which allows users to describe domainspecific discourse styles. And with Cora LiveSpread, users can configure any number of quality assurance (QA) factors.

Automated Normalizer

This feature converts raw, extracted data into a normalized format. Using Cora LiveSpread’s natural language processing (NLP), it matches item labels in raw, extracted data with normalized labels. It also has access to a body of matched labels, organized by type of financial statement and sector.

Exceptions

Cora LiveSpread has proprietary extracting and learning algorithms that go to work on low-quality images. But sometimes document quality is poor and labels are so unusual that it can’t make an accurate extraction or match. Expert financial analysts handle exceptions like these—and once they do, the product remembers the correction for the next time.

Machine learning in Cora LiveSpread

Cora LiveSpread [Figure 3] acquires its knowledge by applying traceable deep learning algorithms based on a domain discourse model. There’s no need to spend time or use expertise to train the machine if all you want is basic automation. On the other hand, if you want more sophisticated spreading, the product can begin learning with a training corpus and labels provided by experts.

Cora LiveSpread applies machine intelligence in two areas. First, it extracts relevant information from documents and normalizes the data to your template of choice. The location and form-agnostic tool discerns forms by deconstructing their physical structure. It has native processing algorithms for PDF text and image documents and can also extract and break up information from footnotes.

Critically, Cora LiveSpread is completely traceable and can adjust to new document structures. Its takes a granular approach to identifying and extracting relevant information from documents. For example, the documents that private companies produce typically
contain financial data and little else. However, for investment banks and public companies, documents often have a lot of extraneous information that makes it hard to identify the right data. Cora LiveSpread addresses such complexity very effectively.

Secondly, Cora LiveSpread’s machine learning categorizes labels into a set of normalized items based on the desired template. It can apply a domain-specific discourse model as a pre-processing step in its learning and processing tasks. In the case of financial statements, accounting identities and hierarchical sections within each type of financial statement can be thought of as the discourse model specific to financial statements. Cora LiveSpread leverages this domain discourse model as needed. For example, if interest expense incorrectly falls under operating expenses in a statement, Cora LiveSpread can warn the user and suggest a correction to non-operating expenses. When an extracted label is misspelled or incomplete, for example, “Accrued Pay,” the tool knows this likely refers to Accrued Payroll or Accrued Payables. It can even provide suggestions based on prior versions of a specific statement or prior statements from that entity or industry. Knowledge of the domain discourse model helps Cora LiveSpread minimize incorrect suggestions.

CASE STUDY

A global financial institution

A global financial institution had Cora LiveSpread up and running in 60 days and now relies on the tool for all of its financial statement processing. The firm added new countries and financial statement formats in a matter of days, too. Cora LiveSpread also configured the financial institution’s extensive rules for treatment of private company statements.

The tool processes over 150,000 financial statements annually from 45,000 companies in 35 countries. International volume expanded steadily, in keeping with the company’s goal, as different countries came on board. The tool automatically uploads spreads to the firm’s internal spreading and credit scoring systems in a fully automated, scalable, and compliant way. As a result, the company has grown its business significantly and shortened the time it takes to make underwriting decisions.

The solution has dramatically improved the timeliness and quality of spreading for this global financial institution, while reducing spreading costs significantly. Figure 4 summarizes the benefits the firm realized with Cora LiveSpread.
Genpact | Whitepaper | 7

**Genpact’s extraction component and AI-based platform are patented**

Figure 3: Cora LiveSpread: Traceable machine learning

<table>
<thead>
<tr>
<th>Area of Change</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compliance</td>
<td>Manual compliance controls. Roll out of compliance changes is complex.</td>
<td>Standards and credit policies institutionalized. One-offs handled as per policy. Risk reviews based on reliable, on-time data. Automated controls. Click-back to exact extraction location in source documents.</td>
</tr>
<tr>
<td>Cost</td>
<td>Very high and reliant on full-time equivalents (FTEs).</td>
<td>Over 70% cost reduction.</td>
</tr>
<tr>
<td>Speed</td>
<td>Best case turn-around was 30 minutes to more than two hours.</td>
<td>80% of financial statements are instantly processed.</td>
</tr>
<tr>
<td>Flexibility</td>
<td>Roll out of changes required extensive training, monitoring, and new controls.</td>
<td>Major changes roll out in hours with minimal or no training. Complete insulation from frequently changing source file formats.</td>
</tr>
<tr>
<td>Scalability</td>
<td>Scalability depended on adding people, overhead, and additional controls.</td>
<td>Fully scalable operations. Volume can double instantly.</td>
</tr>
<tr>
<td>Quality</td>
<td>Potential for inconsistent treatment of statement items across analysts.</td>
<td>Consistency and institutionalization of the process. Implement financial services firms-specific normalization rules, as well as country-specific GAAP treatments, easily and rapidly with no programming.</td>
</tr>
<tr>
<td>Operating model</td>
<td>Typical challenges of a high-volume manual operation.</td>
<td>No peak staffing, attrition, training, quality, and inconsistency challenges.</td>
</tr>
</tbody>
</table>

* These steps include flexible automated QA

**Figure 4: Benefits realized from using Cora LiveSpread**
Summary

This paper showed how a powerful and effective product, based on artificial intelligence and machine learning, can improve financial spreading for financial services firms. The product, Cora LiveSpread, automates all aspects of financial spreading from processing source documents in any format or structure, including images, to the use of machine learning to intelligently learn both in assisted and unassisted modes and non-intrusive integration with downstream systems. We reported results from the deployment of the product in a large global financial institution.

About Genpact

Genpact (NYSE: G) is a global professional services firm that makes business transformation real. We drive digital-led innovation and digitally-enabled intelligent operations for our clients, guided by our experience running thousands of processes for hundreds of Global Fortune 500 companies. We think with design, dream in digital, and solve problems with data and analytics. We obsess over operations and focus on the details - all 78,000+ of us. From New York to New Delhi and more than 20 countries in between, Genpact has the end-to-end expertise to connect every dot, reimagine every process, and reinvent companies’ ways of working. We know that rethinking each step from start to finish will create better business outcomes. Whatever it is, we’ll be there with you - putting data and digital to work to create bold, lasting results - because transformation happens here, at Genpact.com.

For additional information contact, Cora@genpact.com and visit www.genpact.com/what-we-do/digital/digital-products-services/cora-livespread

Follow Genpact on Twitter, Facebook, LinkedIn, and YouTube.