



VAINU

---

# **Automated Data Processing: Building a First-Rate Financial Database of Finnish Companies**

A Vainu Whitepaper

## | Building a First-Rate Financial Database of Finnish Companies

In any market where a company interacts with other companies, there is a constant need for information on their state of affairs. Key data points such as a company's size, its solvency, ownership, growth trajectory, past events, and future plans will likely all impact business decisions involving that company. For example, other businesses may use the data to determine whether to make a sales call, what rate to provide on a loan application, or even to decide whether to do business with that company in the first place.

And as long as that information has been sought after, there has been a struggle finding company information businesses can trust—a reliable database with enough coverage, the highest possible accuracy, and with data that's freshly updated.

One dataset that's seen as especially critical is financial information: at Vainu, for example, over 20% of all database searches include revenue information in their search criteria. And yet while financial statements are public information in Finland, available for anyone to access through the local authority (Finnish Patent and Registration Office, PRH), database providers could still deliver better results in terms of coverage, accuracy, and freshness when it comes to this critical information.

The purpose of this article is to present Vainu's automated pipeline to process incoming financial statements and turn them into structured, searchable information available in its company database in Finland, as to best serve its Finnish clientele.



## **Non-structured data as a hindrance**

In Finland, a key challenge in processing financial information is how this information is filed with the government: a financial statement can be filed in paper, and is distributed externally in a non-structured format—as images.

What this means is that as you can't copy-paste data from an image, collecting revenue data and turning into a searchable database requires opening every statement, finding the desired information, and then writing the numbers from that image down.

To put the problem into perspective: in 2018, a total of 180,000 financial statements were registered by limited companies. Collecting relevant information by hand, with a prompt pace of three minutes per statement, a human would spend 5,5 years on the task—and would still be missing all other valuable information from the financial statement, including the text content available.

And thus any reliable data collection method requires the technical capability to understand images, and use the ability to turn everything in financial statements into meaningful data points in a database.

## **Building an automated pipeline powered by AI**

To arm its clientele with the best possible Finnish financials—fresh, accurate data with the highest possible coverage—Vainu has developed

## | Building a First-Rate Financial Database of Finnish Companies

a highly automated pipeline that quickly converts non-structured financial documents into useful, searchable data points in its company database.

Whenever Vainu receives a newly registered financial statement from the local authority (PRH), it is instantly run through an image recognition system that turns the image's pixels into machine-readable characters and numbers. In most cases, this process is simple—using careful logics and validations, the numbers are interpreted, e.g. identified as revenue or long-term receivables, and stored in the database.

The pipeline also uses artificial intelligence to separate letters and numbers from other elements, and as a result, Vainu uniquely has a database of textual data from the Finnish financial statements, allowing users to find companies in novel ways. For example, a keyword search for “currency” or “foreign” would be an excellent first step in listing companies that do international business.

Since the official format of a financial statement is only very loosely dictated by the authorities, the algorithm must adapt to and understand documents that are radically different in appearance. In some cases, the layout is simply too strange for the algorithm to understand. Here are two real-life examples in images 1 and 2:



In any circumstance where the algorithm can't understand the document, the financial statement will be sent automatically to a team trained in picking out the desired data points. These manually added figures will go through the same data validation process (more on this later), in order to avoid human error.

### **Fresh, accurate information with the highest possible coverage**

To confirm its reliability for Vainu customers, the performance of the pipeline was reviewed across three key areas: the coverage of financial data for all Finnish business entities, the accuracy of that financial data, and the freshness of the data, in relation to when it was made available by the local authorities. As testing performance across all data points made available in a financial statement would be challenging, the review was done by looking at revenue information, due to its prevalence in Vainu database searches made by customers.

Below are the results for each key area, in the order listed above.

#### **Coverage**

In the first test, the percentage of companies with financial figures found in the database was compared to the total number of figures available. The fiscal years ending in 2017 and 2018 were included in the review, as all the financial statements for those years have been registered by the time of testing.

## Building a First-Rate Financial Database of Finnish Companies

It's important to point out that not all financial statements in Finland included revenue information. To get the total number of statements including an announced revenue figure—to compare against the number of statements where Vainu's pipeline had successfully found one—results were validated both manually, and by comparing against other data sources available.

The results are reported in Image 3:

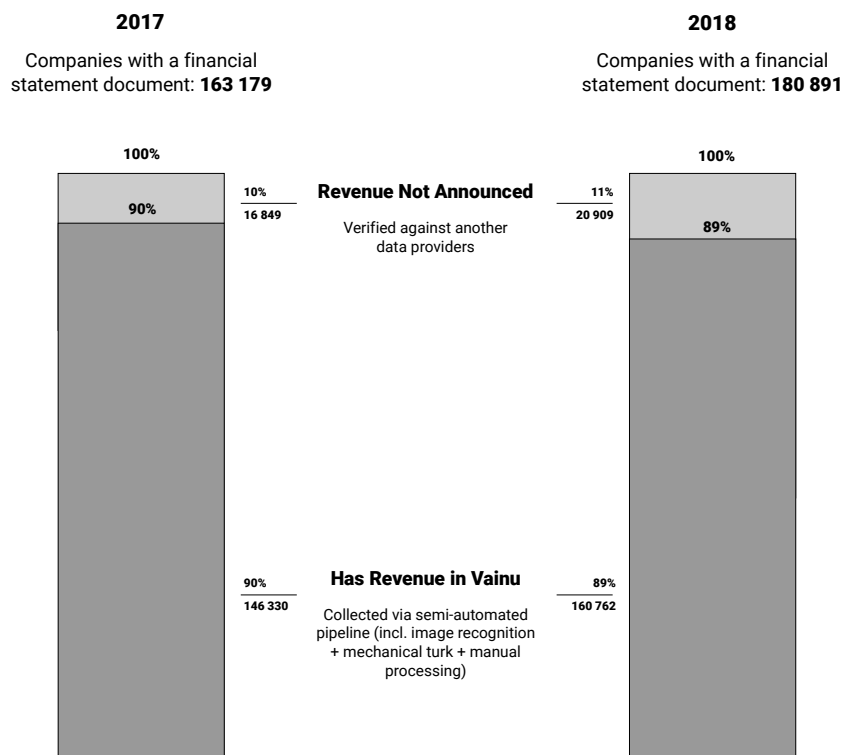


Image 3. Vainu coverage of revenue information in 2017 and 2018

Based on the test results, all 100% of the companies that had revenue information available had results also found in Vainu's database.

## **Accuracy**

In the second focus area, the question was simply how large a portion of the revenue figures in Vainu's database were correct.

The problem with this question is that it assumes there are correct answers available to compare with. And since Vainu's whole data pipeline was born out of the lack of a reliable and accurate database, this comparison cannot be done.

Instead of determining the true accuracy of the database, it's thus more helpful to review the processes in place to ensure the pipeline provides accurate results. Two of the key measures in Vainu's pipeline include:

**Anomaly detection.** To evaluate its feasibility, every financial figure is reviewed against various data points: for example, the previous year's revenue for the company.

**Cross-validation.** Random samples are automatically compared with other open data sources, to ensure the accuracy of the process.

Whenever an anomaly or unexpected difference occurs in any data point, the result is sent automatically for manual review, thus minimizing the possibility of incorrect data in the system.

## **Freshness**

Lastly, when testing for freshness of the financial data collected in



## Building a First-Rate Financial Database of Finnish Companies

Vainu's pipeline, the objective is to make newly found financial figures available for database users as quickly as possible. This can simply be measured by tracking the time taken from the document being made available by local authorities, to the financial data being found in the database.

As the challenge is to understand what processing time would be considered fast and what would be considered slow, a comparison was made with a leading financial database provider in Finland to determine relative freshness in the market. On March 28, 2020 at 5PM EET, a comparison was run to see which database had more revenue information available out of all financial statements published during January 1, 2020 and March 27, 2020. The results are presented in image 4:

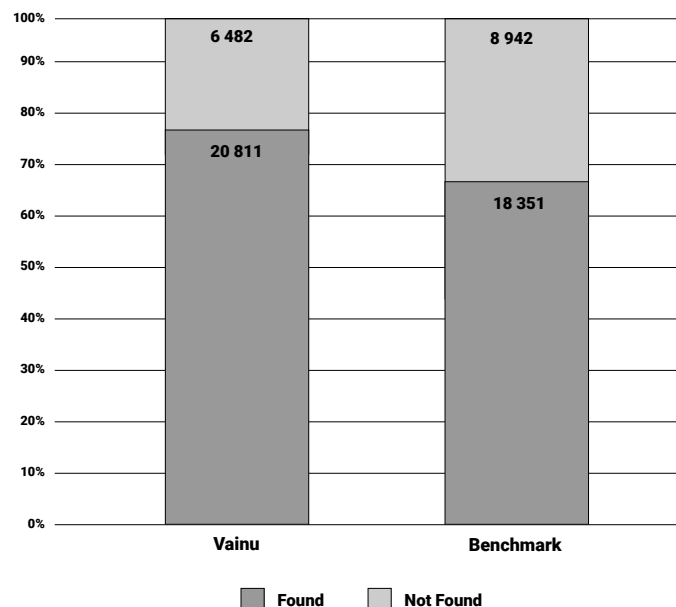


Image 4. Comparison of revenue information for 2020 financial statements

## Building a First-Rate Financial Database of Finnish Companies

As seen in the results, Vainu had successfully collected revenue information for 76% of all financial statements, compared to 67% with the benchmark company.

To further test the speed of the pipeline, another test was conducted as well, now to see which database had more revenue information available out of the financial statements published just the previous day, March 27 (Image 5):

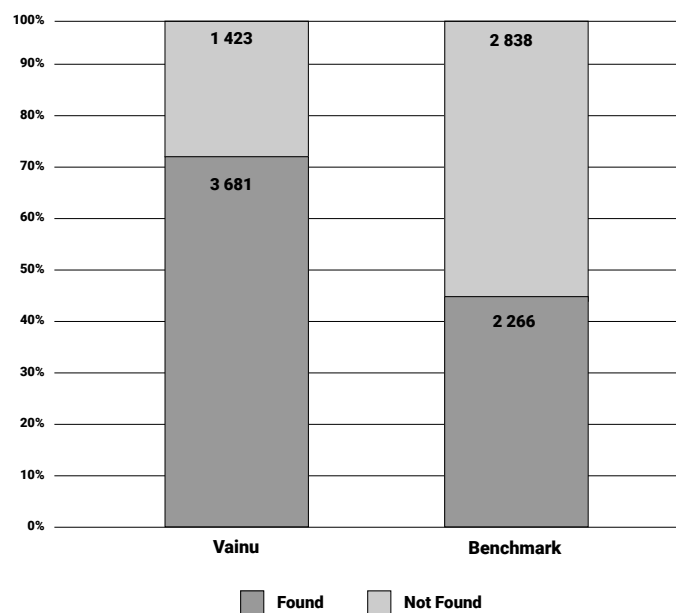


Image 5. Comparison of revenue information for previous day's financial statements

In this test, the difference in freshness is more starkly visible, as Vainu's database had information already available for 72% of the cases, compared to only 44% with the benchmark.

## Being first makes a difference

In a world where data is more and more ubiquitous, the ability to use the available information increases in its importance. And in the competitive landscape we live in, the companies using fresh, reliable data to guide their decision-making processes will come out on top. Being the first one making a move makes a big difference, and trusting automation and the latest technology enables you to be first.



**VAINU**

Vainu is building a Sales Intelligence platform that helps salespeople and marketers move forward. Powered by technology to collect, read and understand all company information ever written, Vainu makes these real-time company insights easily consumable directly in its customers' existing business systems.

Headquartered in Helsinki, Finland, the company launched in 2013 and is now at €15 million in annual revenue with a team of 150, spread across Europe. Over 2,000 sales and marketing teams globally use Vainu's data to personalize customer interactions at scale—ultimately leading to more sales.

Learn more at [www.vainu.com](http://www.vainu.com).