

How Ocrolus CEO Sam Bobley built his document analytics company into a \$100 million fintech firm

We spoke with Sam Bobley about how he built up Ocrolus, and why the company plans to expand into banking and mortgage document reading.



IVAN DE LUCE



10.9.20 10:00 AM IN **PEOPLE - SCALING UP**



If you've ever answered CAPTCHA questions to prove you're not a robot, you were probably unwittingly helping Google analyze print documents to turn them into digital ones. [Ocrolus](#), one of the leading document analysis companies in financial services, uses similar tech to digitize financial documents.

Sam Bobley, cofounder and CEO of New York-based Ocrolus, took that idea, which started in his parents' kitchen in 2014, to a reality. Today, Ocrolus goes through nearly 500 million financial transactions per year, turning over 30 million pages of financial documents into data with 99% accuracy. Inc. Magazine has named Ocrolus

Thinknum: You started Ocrolus from your parents' kitchen. How did the idea for the company come about?

Bobley: I was trained since I was a little kid to be an entrepreneur. My dad is a serial entrepreneur, he started a bunch of companies throughout his career. And when I was growing up, he was building and scaling a financial technology company, a payment company called Phonecharge. He ended up selling Phonecharge in 2006. And that was when I was in high school. So my early childhood was watching my dad scale and ultimately sell a payments company.

I was naturally interested in the financial technology space and wanted to follow in my dad's footsteps. And the genesis of Ocrolus came from a conversation I was having with my dad when I was in college. He was telling me about a meeting he had with his elder law attorney. And the elder law attorney was complaining about having to review hundreds or thousands of pages of bank statements and other financial documents for every Medicaid application. For every Medicaid application, when someone applies for long-term care Medicaid to enter a nursing home or assisted living facility, they're federally required to submit 60 months' worth of financial information along with their application. And what happens typically is elder law attorneys or nursing homes review these documents and find noteworthy transactions. For example, any transaction above a thousand dollars he proceeds to make notes and indicates what that transaction is for. I knew nothing about document analysis or Medicaid or any of that stuff at the time, but what I did realize was that this was a problem. Why are high-paid lawyers and paralegals

“Why are high-paid lawyers and paralegals spending their day, page by page, line by line, reviewing bank statements with a highlighter in a super old-fashioned way? It just didn’t make sense” - Bobley

It really started as a research assignment. I reached out to a buddy of mine from high school, we were in my parents’ kitchen, and we were researching this problem. We were trying to better understand why you just digitally retrieve all the information directly from a financial institution. Why can’t you use technology to extract all the characters and text off of the documents automatically. And why are so many people doing this job manually? Through our research, we came to the realization that there was no technology on the market to do this in a turnkey way, and to handle all of the different types of documents that one might see.

What was the real breakthrough moment for Ocrolus?

What we realized in the research was the existing ways of doing this created too many errors. There was technology out there that could do the job, let’s say 60 or 70 or 80% effective, but there was no technology that could do it perfectly. And the reason for that is because financial documents come in all sorts of different formats. They can come as edocuments, faxes, scans, cell phone images — bank

The lightbulb moment for us was to piggyback off of Google's use of CAPTCHA technology. So I'm sure you've gone to a website and it says to you, "Prove you're not a robot and type in these digits." What many people don't realize is you may actually be helping Google transcribe millions of old books, menus, and magazines into digital data. Google tried to use OCR (optical character recognition) to read and archive millions of documents, but they ran into accuracy problems for the same reasons I'm describing. Different formats, fonts, poor image quality, they couldn't read every word every word accurately.

So what they did was quite brilliant. They snipped the words they couldn't read with a high degree of confidence, and they routed them to people like me and you to validate them in CAPTCHA tasks, effectively leveraging us to perform quality control. And they introduced all sorts of interesting algorithmic checks as part of this. For example, they route the same task to a statistically significant number of workers. Sometimes you'll see two words in a CAPTCHA box. One word is a control word that they have the answer to, the other is a snippet from an old magazine that they need your help validating. Long story short, 30 million-plus books, menus, and magazines are now searchable online in terms of accuracy, and simultaneously Google was able to take that data and use it as training data to teach their system to more effectively read weird fonts and formats and poor quality images on future passes. Google's now doing the same thing for autonomous vehicle training. It's the same concept. Whatever image their system couldn't auto-classify, humans classify it, we create a perfect dataset, they use that perfect dataset to train their models to do it automatically.

whatever we can with a fully automated machine approach. Whatever we can't read automatically we splice into smaller tasks, we have our own employees verify the information, we then layer in all sorts of algorithmic quality control checks to make sure our workers did the work accurately. We return perfect data to our customer, and then we use the feedback loop for the data to continually train our models to become smarter, thus mitigating the need for human review on future passes. With that model, we've been able to create a really unique category of document analysis, where we can turn anything into perfect data, and really solve the problem for our customers in a turnkey way.



What were the first steps you took to get Oculi off the ground?

at that time we'd moved into a very small office in Glen Cove, New York, on Long Island. And I was 22, my business partner was 22, [we] hired a couple of guys who were from Hofstra University, a local college, to help us build the first version of the platform. Simultaneously, we started talking to beta customers, people who would test it out and give us feedback. Originally we were talking to lawyers and accountants, people focused on this Medicaid type use case, that was really the genesis.

As we got the product ready for testing, got some people lined up and got their initial feedback, that's when we realized we had something that was more than an idea and really a big business. And specifically it was the customer experience. When customers tested the product, and they were used to testing products that got them 75% accuracy, and then they tested ours and they got perfect data, they were like, "Oh my God, this is amazing, we could pay for this." Even before we started charging customers, one person who was testing the software said to us, "Where do I send my credit card information, because I'm ready to pay." We weren't even ready to launch, but they were ready to pay. That was kind of the indication, like, "Hey, this thing is ready for prime time, and we have to scale up."

"We weren't even ready to launch, but they were ready to pay. That was kind of the indication, like, 'Hey, this thing is ready for prime time, and we have to scale up'" - Boblev

the goal was to attract more senior talent. A team is easier to recruit in New York City than on Long Island. And we did something else really critical. We pivoted and remained very focused. Even though our initial use case was for lawyers — and we were actually very successful with that, in our first year of launching the product, we had dozens and dozens of lawyers using the product — what we realized in that first year was that lending was a bigger market opportunity for us.

We got introduced to a small business lender who had much more volume, in terms of documents, than your average lawyer, and who really seemed like a perfect early adopter customer for us. We made that pivot, which was a risk for us at the time, to focus exclusively on this lending market, and to really develop lending-specific product features. For example, fraud detection and analytics, and other things that can wrap around our document processing engine, to make the product even more valuable to lenders. That decision to focus on lenders really led to our rapid growth. We were then able to attract investors, who had backed many of the leading lenders. We were able to get case studies out the door with a few early adopters, highlighting how we increased the accuracy and speed in their back office processes, and we were able to do a lot of things to create the snowball effect, and ultimately lead to us becoming the number one fastest-growing fintech company between 2015 and 2019.

So Ocrolus was named the fastest growing fintech startup by Inc. magazine. Did you expect it to grow so quickly when you founded the company?

I think it was a domino effect. We knew the business was going to be big, but I think

customers.

“It’s really painful and hard to get the first couple of big customers in, and once we got a couple in, it was like we blinked, and we were over a hundred customers” - Bobley

If I had to sum it up, it was like realizing the product-market fit. You think you understand it from a conceptual point of view, like what is product-market fit, but then you actually have the moment where things just click. You have a small sales team, but then all sorts of inbound customers are coming in, there’s no friction in the sales process. It was just an amazing run we had over the last couple of years, but I think more importantly, it painted the picture of the potential for us in additional markets. We were in online lending, which is a very small niche, its own universe. The bigger world for us was mortgage lending, banks, legacy financial institutions, maybe one day health and insurance, and all these other areas where our technology theoretically could be applied. Today we are still focused on lending. We are moving from online lenders to more big lenders and banks. But I think the growth over the last few years has given us the confidence to continue investing, to really focus on replicating our success on a larger scale with legacy financial institutions.

banks and legacy financial institutions have no choice. They have to move their processes online, and they have to invest in automation. And what they're doing as part of that is they're looking to the big online lenders. They're looking to the PayPals and Squares and SoFis and LendingClubs of the world and trying to figure out how these guys were able to move everything online and make their processes more efficient. One of the reasons is because they have OcroLus in their backend powering their underwriting process. We're one of the tools that really help the online lenders, and I think we're now in a position where the banks can leverage our technology to increase efficiency in their own back office. I saw a stat the other day, and it said that 1% of the loans in the world are made online. Obviously this is a pre-COVID stat, but I think it paints the picture. We had been successful in that 1% bucket. Our objective now is to back that 99% that are offline, and help bridge them so that they can get online and modernize their process. I think it's the right place and right time for us to be leaning into this opportunity.

How has the pandemic affected OcroLus?

It's helped us increase velocity in terms of new sales opportunities, and it's helped us gain awareness with mortgage lenders and banks who are eager to invest in process automation, and online lending. It's the classic, "Never let a good crisis go to waste." We've tried to understand the effects of the pandemic, and how it will cause changes in the landscape of financial services. I think our product is one of many in this layer of fintech infrastructure. Companies like Plaid that are really revolutionizing the way financial services companies operate. We are one of those companies, with a focus on document analysis.

I think so. I think the next 18-24 months are critical in terms of planning, like how financial services companies plan to modernize their back offices. I do think the world will return to somewhat of a normal state. I still think banks will have branch locations, and not all in-person interactions are going to be eliminated, but what I think we'll see is that financial services will have a multi-pronged channel strategy. Historically, a lot of these legacy lender and banks — your only option was to come into the facility, work with a loan officer, hand them a packet of information, and do it the old-fashioned way. I think what we'll see is almost everyone will offer a dual path, where you can still come in and have the personal touch, but you can also have an online-only option where you can do the process digitally sitting at home on your laptop.

Were there any initial challenges at the beginning? Any bumps in the road growing the business?

One of the biggest challenges in enterprise software, specifically in the financial services space, is security and compliance. A lot of the bigger lenders and banks want to make sure that you have various certifications to ensure data is handled with the utmost protection. That's hard for an early stage company. It's hard to get ahead of that. It's something that we had to invest in, and we've done a great job investing in that over the last couple of years. We brought on a head of compliance, a head of cybersecurity, we've now completed our SOC 2 type 1 and type 2 certifications. But nevertheless, it's a big challenge. It's the type of thing that takes a lot of time and a lot of energy to get right. There's no silver bullet. I would just caution other entrepreneurs, if your plan is to go whale hunting and go knock down

“A lot of the bigger lenders and banks want to make sure that you have various certifications to ensure data is handled with the utmost protection. That’s hard for an early stage company” - Boley

Startups, when you’re first getting started, you don’t have the credentials to say, “Yes, I’ve serviced these other companies, this is why you guys should use us.” And you need early adopter people who are willing to take a chance at being one of the first movers. And something that gives them peace of mind is being buttoned up from a cybersecurity and compliance point of view. We did the right things to invest — if I had to do it all over again, I would do the same thing, but I would do it even earlier in our company’s life cycle. I think it’s important to do that really early in your journey if you want to sell to banks.

What's next for Ocrolus? Any new industries you're taking on?

The biggest one for us is mortgage. Most of our large customers are small business and consumer lenders. And what we do is we make the underwriting process faster and more efficient. [Mortgage] is probably the biggest market opportunity for Ocrolus. The average mortgage application contains hundreds of pages’ worth of documents. It’s a very slow and inefficient process today. Most of it is done

really compelling market opportunity to help these mortgage lenders modernize and begin to take their processes online.

What advice would you give to young entrepreneurs out there?

I think the number one piece of advice is persistence. In the early days it's extremely challenging. Along our journey, we had all sorts of mistakes and hurdles to get to the inflection point. We pitched hundreds — literally hundreds — of investors who said no to us, didn't respond to emails. It's easy to get demoralized, to want to give up, because it's tough.

“One of the quotes I like to share, it's a Nelson Mandela quote, which is, ‘I never lose, I either win or learn’” - Bobley

One of the quotes I like to share, it's a Nelson Mandela quote, which is, “I never lose, I either win or learn.” So if somebody says no to you, take it as a learning opportunity. Don't get dejected, don't get too down on yourself. I think that kind of mentality is super powerful in the early days. If you really have conviction in your idea, and you know it's going to be big one day, you have to just power through the early days. There's an inflection point. There's one moment where things start working, and you're on the uphill, and it's really awesome from a founder's perspective to see that shift, and to see investors courting you instead of you

FOLLOW IVAN DE LUCE ON [TWITTER](#) AND [LINKEDIN](#)

SIGN UP FOR OUR NEWSLETTER

Start your day off with our weekly digest.

Join

Latest Articles

Who really works for

OpenSea?

Here's what quirks in LinkedIn headcount can tell us about crypto's impact on the job market.



CHRISTIE ... 7.22.22 11:41 AM



Twitter's endgame in its case against Elon Musk

The social media platform is trying to force Musk to complete his takeover deal. Is that really what it wants?



CHRISTIE... 7.14.22 10:56 AM





Bolt and client ABG bury the hatchet and become “proud partners” again

ABG, owner of Forever 21 and Lucky Brand, said a lot of nasty things about the one-click checkout star in a lawsuit earlier this year. Now that’s all water under the bridge.



CHRISTIE S...

7.7.22 9:20 AM

OUTCASTS



A brief history of why Martin Shkreli got banned from Twitter

The following is an excerpt from SMIRK, a memoir of journalist Christie Smythe’s unusual



Recession and bankruptcy fears are rising, corporate filings show

Mentions of “recession,” “bankruptcy,” and “Insolvency” in SEC filings have all been steeply higher over the past two years compared with 2018 and 2019.



CHRISTIE S... 7.6.22 6:23 AM



Analysis: the sharks are circling crypto, and it's going to take a lot to fight them off

A lot of bad things are happening in crypto now. And it could get a lot worse.



CHRISTIE S... 7.5.22 4:52 PM



This is an edited excerpt of serialized memoir SMIRK, describing journalist Christie Smythe's relationship with the notorious former drug CEO, who was recently released from prison.



CHRISTIE ... 6.27.22 6:58 AM

FEATURES



"Repulsive," "reptilian," "diseased": What John D. Rockefeller had in common with Martin Shkreli

The following is an excerpt from SMIRK, a memoir of journalist Christie Smythe's unusual relationship with "Pharma Bro" Martin Shkreli. It should be viewed as opinion. You can read

[more at www.smirkbook.com](https://www.smirkbook.com)

View All

Thinknum

For Customers

Product Manual

Product Updates

API Documentation

Company

About

Explore Datasets

Resources

Careers

Press Kit

Solutions

Thinknum Alternative Data

KgBase

Industry Products

Investor Intelligence

Business Intelligence

Education

Media

About

Careers

Revisions & Errata

Newsletter

External Data Conference

Get Featured

RSS

Media Services

Content Studio

Advertising