



EXPERT INTERVIEW

UPDATED
01/05/2023

Tony Xiao, founder and CEO of Venice, on the opportunities in financial data aggregation

TEAM

Jan-Erik Asplund
Co-Founder
jan@sacra.com

DISCLAIMERS

This report is for information purposes only and is not to be used or considered as an offer or the solicitation of an offer to sell or to buy or subscribe for securities or other financial instruments. Nothing in this report constitutes investment, legal, accounting or tax advice or a representation that any investment or strategy is suitable or appropriate to your individual circumstances or otherwise constitutes a personal trade recommendation to you.

This research report has been prepared solely by Sacra and should not be considered a product of any person or entity that makes such report available, if any.

Information and opinions presented in the sections of the report were obtained or derived from sources Sacra believes are reliable, but Sacra makes no representation as to their accuracy or completeness. Past performance should not be taken as an indication or guarantee of future performance, and no representation or warranty, express or implied, is made regarding future performance. Information, opinions and estimates contained in this report reflect a determination at its original date of publication by Sacra and are subject to change without notice.

Sacra accepts no liability for loss arising from the use of the material presented in this report, except that this exclusion of liability does not apply to the extent that liability arises under specific statutes or regulations applicable to Sacra. Sacra may have issued, and may in the future issue, other reports that are inconsistent with, and reach different conclusions from, the information presented in this report. Those reports reflect different assumptions, views and analytical methods of the analysts who prepared them and Sacra is under no obligation to ensure that such other reports are brought to the attention of any recipient of this report.

All rights reserved. All material presented in this report, unless specifically indicated otherwise is under copyright to Sacra. Sacra reserves any and all intellectual property rights in the report. All trademarks, service marks and logos used in this report are trademarks or service marks or registered trademarks or service marks of Sacra. Any modification, copying, displaying, distributing, transmitting, publishing, licensing, creating derivative works from, or selling any report is strictly prohibited. None of the material, nor its content, nor any copy of it, may be altered in any way, transmitted to, copied or distributed to any other party, without the prior express written permission of Sacra. Any unauthorized duplication, redistribution or disclosure of this report will result in prosecution.

Published on Jan 05th, 2023

Tony Xiao, founder and CEO of Venice, on the opportunities in financial data aggregation

By Rohit Kaul



EXPERT INTERVIEW

**Tony
Xiao**

Founder & CEO
Venice



Background

Tony Xiao is the founder and CEO of open source aggregator of aggregators of financial data Venice. We talked to Tony to learn more about the competition in the financial data aggregation space, emerging trends in financial data aggregation, and the need for a Segment-like layer in banking data stack.

Interview

What core problem were you trying to solve with your last project Alka?

Alka was accounting for individuals with complex finances. Usually, you only do accounting when you are a business. As an individual, you don't generally do it. The idea of Alka is to reduce the friction and make the cost of accounting sufficiently



low that it can actually make sense for individuals whose personal finances are fairly complicated.

On the high end of the spectrum are the family offices of the world running actual corporations. On the other end of the spectrum are people who are more like myself, and just happen to lead a more complicated life. I have assets in a couple of different countries and owning a business or sharing expenses can get a little more complicated. That was where Alka started.

Think of it as mint plus superhuman for the people who really need a higher tier experience to support more capabilities in their financial management.

However, it didn't really work as a standalone product in that particular format. What we realized was that people bought into the idea much more than the effort they actually put in to make the product work. That also reflected in that gap in the product. Perhaps, we were a little too idealistic when it came to wanting to have the perfect accounting records for everyone. We really missed the ballpark on the 80/20 rule, and it just took practice to get the product to work and cover all the edge cases.

We learned why people have professionals manage this because we just didn't get the product to a point where the friction was sufficiently low that it actually made sense for people to invest the time in doing it, regularly. The value that we were giving them was less than the amount of effort required.

We had a bunch of people paying us. But, it was one thing to pay for a product because you believe in this idea, "Wouldn't it be great if I had a perfect set of... If I could run my personal life like a business and have a perfect set of books?" It's another to realize, in retrospect, that it's more like a gym membership where it's much easier to pay for a gym, at least for a certain group of people, than it is to go to the gym every day and do the workout. That's the real hard part.

Gym membership is easy to pay for, but somehow, gyms make their business model work. We could perhaps have resolved this but in any case, it didn't really make sense for us, for our kind of product. We didn't see the engagement in the product



that we really wanted though yes, the idea was to build a superhuman for personal finance.

What did Alka's data aggregation stack look like?

We had the standard stuff—Plaid, Yodlee, and Salt Edge for covering some international markets. We built integration with this aggregator which was doing a sort of Plaid for crypto—Zabo—but they went out of business. We also built a mechanism to import CSV files that are in bank-specific CSV formats—Capital One, for instance, or First Republic.

Sometimes, when you're talking about accounting, you want to create all kinds of ways for people to get data and there must be a way for this in the system even if it's not supported by an aggregator.

We developed a CSV import process where people could just select their institution, their bank. We already pre-prepared how to import data from that specific institution so they didn't need to worry about formatting it in a particular way, to fit our process. Of course, we also had a general purpose way to import data—manual entry—that was always a component.

We also built custom integrations with a few institutions that just weren't supported by any aggregator, including Venmo, for example, at one point. We reverse engineered a couple of API ourselves in order to get data from them.

Why did you use both Plaid and Yodlee when they cover essentially the same types of banks?

It was ultimately coverage. Most of the institutions that these aggregators claim are overlapping.

Yodlee was one of the only companies that—with help from their team in India—maintained these aggregations and screen scrapers. It's really expensive to build that and it's not very profitable because you have a long tail of institutions for which you need to maintain pipelines.

Plaid didn't actually develop most of them, themselves. They may be built 500 or a thousand out of 19,000 and the rest they got from other data aggregators.



So, when you see these aggregators on the market, 90 percent of the banks they support are the same because they bought it from this infrastructure player and just licensed them under different brands. It's not like when you have Plaid and Yodlee, you'll double the coverage. The idea is that they're all just rebranding the same set of underlying connectors from this company that sells them.

This is why, when you're looking at a Venmo or Revolut—non-traditional financial institutions—many of these fintech apps where people increasingly have money in, you'll find very spotty support for them because there's actually, relatively much less development of data connectivity than it will seem based on the number of aggregators out there. That's because they share most of them and they have the top connection.

The pattern I see is that each aggregator will maintain the top 10-100 connections themselves and then outsource the rest to the same underlying infrastructure player. That answers, in my view, why they are more overlapped than not.

At the same time, they do have slightly different focuses. We wanted to get as much coverage as possible, so we just kept on running into financial institutions not being supported. It would be a lot less work for us if one aggregator just did it all, but that does not exist.

The question then is, what should we do as the fintech app?

We were focused on giving people a single view of all of their financials, so we didn't really have a good choice, especially in our target segment where people had a lot of different assets and were all over the place to be able to bring it together. We just needed to have additional data sources that weren't supported. No one aggregator could do it for us.

There are literally banks that you can search in Plaid and you will not find. For example, Venmo, it just won't be there. It's one of the most popular apps in the United States. Or Cash App—they simply won't be there, so if you want to support them, the only choice is to use something else that supports them or build it yourself.

You can think of it as a fail-safe, but for certain institutions, the failure rate will be a 100 percent because it's just not there.



We see that fintechs like Venmo use only one data aggregator while you mentioned using both Plaid and Yodlee. Do you see any trade-off in terms of using one vs. using multiple aggregators?

It's one of these situations where there's a lot of vendor lock-in in the ecosystem. It's inherent, and from both, a user experience perspective as well as technical perspective. When you speak to the many companies that have built their technical abstractions around Plaid, inevitably they face challenges once they get to a certain scale. But by that point, it becomes really expensive to change the architecture. If you didn't think of multiple aggregators from the start, then the cost of changing is actually very substantial because you need to build this multiplexer layer and think about normalizing, standardizing between them.

From a user experience perspective, some of the aggregators, Plaid, for instance may not share user credentials like banking username and password, with you as the fintech. If you want to use a different aggregator, all of your users will have to re-authenticate and that implies bad user experience. It's also very expensive from an engineering perspective.

In terms of the benefits, it really depends. Venmo, for example, is only operating in the United States. So a lot of the concerns that we had aren't really a concern for Venmo, because Plaid actually has pretty good coverage in the United States. The type of places that Venmo needs to connect to, are basically traditional banks that you ACH money in and out of Venmo. So Venmo doesn't really care to connect with other fintech apps per se, because that's not their purpose.

Venmo isn't trying to build a personal finance dashboard. The only purpose of doing bank authentication is to be able to send money around—ACH, that's it. For Venmo, it's less painful, there's less incentive to migrate, and it'll also be extremely expensive, so I don't really see them doing it anytime soon.

Square had a project to get on multiple aggregators for years, but it was just hard for them to really make it happen because of similar reasons that I was just mentioning for Venmo. There are still a lot of incentives, for example, when it comes to contract negotiation—the more you're locked in, the worse pricing you're going to get. It's just the natural law of how it



works—the incentive structure. So, there is definitely a reason not to do it, but again, it's just the cost-benefit analysis.

What then, differentiates a Plaid from a Yodlee or MX? You mentioned some very specific institutions that Plaid has and Yodlee doesn't. But apart from that, is there anything in terms of data connectivity or quality of data?

So when Plaid entered the scene, the biggest differentiator for them was the developer experience and the business model. Pre-Plaid, you needed to sign a contract with a minimum of at least \$500 a month. You couldn't just get access to an API or sign up for an API key. You'd have to talk to a salesperson.

Then, Plaid came onto the scene and said, "Hey developer, you can get started in two minutes for \$0. We're going to create a much better developer experience around the documentation, around just the APIs than Yodlee did." That's what really differentiated Plaid from Yodlee, and what allowed Plaid to get themselves onto the scene.

Yodlee has been playing catch up since. Now, you can actually get a Yodlee plan for a much lower cost as well but, you can still see the level of difference in API documentation and ease of use, for sure. It's hard to build that and it's a cultural thing, but they of course have been playing catch up in many areas.

With regard to data quality, I've heard things both ways. In our personal experience, it's really institution specific. It's hard to just say in general, Plaid has better data quality versus Yodlee or vice versa. What's a lot easier to tell is if a specific bank is supported or not. It gets harder the deeper you want to go, "Well, for what percentage of the time is it supported?" "We say it's supported but will it fail 100 percent of the time?" That, you can only know by trying.

Technically, Capital One was on Plaid's list of supported banks for a long time, but for an entire year and a half, it was pretty much failing 99 percent of the time. And then, the deeper you go, "What about the quality of the transactions themselves? How many months or years of history do you get back? What about the richness of that data?" The more you want to go into it, the harder it is to actually evaluate. By the time you get a few levels down, it is really hard to just say without actually running real data and real traffic against them, to see which



one actually works better for one specific use case. It's quite challenging to say that in the abstract.

That being said, with open banking becoming more and more adopted, I do see disparity in data quality becoming less and less of an issue because if all of these vendors are going to be talking to the same underlying banking APIs that the banks are coming out with, then, the data will actually just be the same.

It's only when you're scraping that there's a bigger concern around what is the quality of the scraper you built because once it becomes standardized APIs, then, the quality should be much more standardized as well.

Now there's a whole other set of data activity called data enrichment, which is not the data that comes from the banks themselves, but the additional processing that the data aggregators do on top of them. I'm seeing an interesting trend where there are third-party players now, like Heron Data and others, whose entire purpose is to enrich data. They don't aggregate any data, they're breaking the aggregator apart saying, "We're not going to do the job of aggregating data. We're just going to do the job of enriching data. You need to give us data from wherever you have it."

When you say data enrichment, do you imply that they are combining data from different sources or are they splitting this in certain unique ways?

If you look at most of the transactions on a bank statement in the U.S., it's a bunch of gibberish with random letters and numbers. Some of this makes more sense than others, but it's not very clean.

What enrichment does is, clean up all of the random numbers and digits that a consumer is not going to recognize or care most of the time and then show, for example, an actual name of the merchant—Walmart or Amazon instead—with a logo, and to create a better user experience when you're looking through the list of transactions, to help people recognize their transactions. That's an example of enrichment—taking data and just making it better and more usable for the end user.

Another would be to run some machine learning or just a heuristic algorithm to figure out "What are the recurring subscriptions?" because your bank cannot tell you what your



subscriptions are, because it's just one transaction after another.

But these enrichment players, they can look at it. Plaid also does that now. It's like they are bundling in more and more of these services, going deeper and deeper into the value chain. There are enrichment companies who can help you look at that data and tell you “Hey! What are recurring subscriptions?” And then, they can go even deeper, “What's this person's annual income? What's our prediction of their annual income? What's our prediction of their creditworthiness based on the cash flow that we're seeing for them? In the case of a company, let's estimate the cash flow and balance sheet.”

Do you see data enrichment companies replacing data aggregators?

The question is, “Why are you using Plaid in the first place? What's the purpose? What is the job to be done here?” In Venmo's case, the job to be done is, “We just need to transfer some money, we just need an ACH account number and verified bank account. We really don't care about any transactional data” in which case Plaid is enough.

Ultimately your purpose as a lender is to figure out the credit worthiness of the potential borrower. You have raw transaction data, but there's still a pretty large gap between having that and then figuring out credit worthiness. It's about going higher up the value chain.

If you're a personal finance app, then maybe you care about a list of subscriptions that your customers have rather than just a raw transaction. It really depends on why that data was needed in the first place. What is the thing that companies are looking to do with that data?

Aggregators are going into the enrichment space themselves. They're going into, “How do we make this data more actionable? What is the end purpose that our customers are trying to do with this data?” That's because data access is going to become more and more commoditized especially with open banking.

So in order to keep generating more profits, they need to develop new product lines that go more and more up the value stack.



Plaid recently started this entire ACH payment system trying to build a parallel network to Visa and MasterCard. What do you think is prompting their move into ACH space?

If you've seen the Visa versus United States antitrust case, there's actually a lot of information in there. Visa's concern was that Plaid could use this network of financial institutions to cut them and MasterCard from the massive amount of payment volume they've facilitated.

Now that the acquisition wasn't completed, that's just going according to plan. It's pretty clear that Plaid is trying to be an alternative form of payment to credit cards. ACH is already the largest, by volume, I'm sure, the largest but it's not very convenient for day-to-day use. The bet here then is, "How do we make ACH more convenient and safer so that consumers use it more?"

At the same time, ACH is only a U.S. protocol, and there's a lot of headwinds as well. Visa and MasterCard are so entrenched that, for example, when you buy something in the store, Visa and MasterCard take a 3 percent cut. But if you pay by cash, it's not like you get a 3 percent discount. You still pay the same price as somebody who's paying by card, which then earns cash back and builds a credit score. The entire U.S. financial system is basically built around using credit cards.

As a consumer, you are putting yourself at a financial disadvantage by using ACH. There is less guarantee in terms of chargebacks and ACH is slower. Plus Visa and MasterCard have managed to create a system where even if you don't use ACH, the entire U.S. consumer base collectively pays for the additional costs. It's a tax on the whole economy.

So unless there are systemic changes whereby people actually across the board coordinately push away from Visa and MasterCard into the ACH or real time transfers, it will be difficult to see ACH-based consumer payment rails really overtaking Visa and MasterCard anytime in the foreseeable future. The across the board push is going to be pretty difficult given that Visa and MasterCard are the biggest players in the space so they're very entrenched.



Tell us about Venice. What's the inspiration for this?
What's the core problem that you're trying to solve?
What's the value proposition of Venice for a fintech?

It came from building Alka. We were spending 30 percent of our time in engineering, dealing with data integrations and basically helping our customers get their data into the product. As a result of that, it occurred to us that we're probably not the only ones doing that among the fintech companies.

The inspiration is like, "Hey, can we build that out?" It's the infrastructure that we were spending so much time building, so could we turn that into a service infrastructure for other people? It's what we wish we had as we started building.

The vision is to build a single framework for integrating with financial data of any kind and it's designed to sit in between the fintechs and the Plaids of the world as a multiplexer where we help companies get data wherever they're available. Then we provide the bridge to do that, sort of like an aggregator of aggregators.

A similar model would be Segment for customer data. The overall vision is just to reduce the friction of accessing financial data and make it as easy as possible for companies, or even individuals, to get access to their own or their customers' financial data with the least amount of cost and friction possible.

We will help them get that data from Plaid or Yodlee. They just need to build one integration and then they'll get data from anywhere—Plaid or Yodlee or new aggregators that don't exist yet, it will be future proof in that way. And because it's an open source framework, they can also use it to build custom integrations that are specific for their use case, but still leverage the same underlying infrastructure.

So for example, if you wanted to build—in our case it was Venmo—but, if you were using Plaid and you wanted to build a Venmo integration, you'd have to rebuild the entire UI backend infrastructure especially. How do you securely store user banking IDs and passwords? To rebuild all of that from scratch just to add on one single additional integration is a lot of work.



Or, if you wanted to improve upon an integration that currently isn't working well, you're out of luck because you don't have access to the source code. Again, you have to redo it from scratch.

Our goal is to make it so that the infrastructure is just there and people can build their own custom integrations, which is one-hundredth of the work of rebuilding everything from scratch. To just build a single integration is actually quite easy but the entire infrastructure to manage it and the user interface, the security etc., is pretty hard.

We want to create the experience where companies are free to build their own integrations, free to use any of the data aggregators on the market, free to directly plug into, because that's the world we're moving into—financial institutions that have open banking APIs and future proof themselves when it comes to financial data access.

By integrating with Venice, a fintech app gets better coverage, negotiating leverage with vendors because you're less locked in, future proofing, and savings on engineering time.

With Plaid owning the end to end data pipeline in a fintech app, where exactly does Venice sit in the data aggregation stack?

So the first step for a customer in an app usually is choosing which financial institutions they want to connect to. That's where we come in. We have an index of all of the financial institutions supported by all of the data aggregators that we have integrations for. When customers choose a financial institution, they then, based on the rules that you can set as a developer, get routed to Plaid to then complete that experience.

However, if they choose an institution that's not supported by Plaid, then, they will get routed to a different aggregator that supports it. So Venice sits just a step before the Plaid linking experience.

What's the hard part about building a Segment-like aggregator of aggregators for financial data?



There are a couple of startups trying to do something similar and all of them have different takes on the space. Fintech is quite different because you're talking about consumer permissioned data and not just a data infrastructure play.

For example, an end user would never know Segment existed because it works behind the scenes. However, they would have to go through a workflow to connect their bank account with Plaid.

Plus, at the beginning, all of the APIs are closed source, so most of the value comes from reverse engineering those closed source, closed APIs.

Consumer permissioned data is very interesting, because this is very unlike Segment. I recall hearing that MX talks about themselves as being an aggregator of aggregators. Have you come across that positioning?

Absolutely. It's good that they're doing that because I have a list of two, three dozen data aggregators across the world. As someone who lives and travels internationally I have assets in different countries and there is not a single app out there that can aggregate them together. I know I'm maybe a little bit of a special case compared to a lot of people, but my personal vision is for that data access to be really truly seamless and to be effortless, and that's just not the case. I don't see that being the case anytime soon, because by the very nature of the market, each data aggregator has to focus on a specific market, with a specific set of use cases. That's why you see the Plaid for Southeast Asia, the Plaid for Africa, etc.

Venice is open source but we have typically seen aggregators as being more proprietary. What are your thoughts about that?

When you're closed source and proprietary, the problem is that it really puts developers in a helpless position when something isn't working. They can't add a new integration, they can't fix something that's broken. That's my bottom line with it. I want to enable frictionless data access for everybody who needs financial data.

And when you're closed source and proprietary, and that's just a tried and true model so far, it stifles innovation and



community contribution. My guess is they are guarded as one of the most valued secrets—trade secrets—and that enables competitive advantage.

But as the integrations themselves become more and more commoditized, I see the future being a more open source one, where, "Hey, let's go up the value chain" is becoming more and more interesting.

The only way I see frictionless financial data working is when everyone can contribute to enabling that. Have a bank that's not supported? Just build a custom integration and you can leverage the whole ecosystem. So that's the vision for why it's open source.

With regard to Venice, what kind of traction are you seeing? Any early thoughts on user acquisition and the next steps?

We're still pretty early, so just testing privately with a few companies at the moment, building more based on what they need. The next step for us is to actually get this launched.

One of the most interesting use cases is of companies wanting to use our product as decentralized on/off-ramps for crypto. If we're going to move towards a global financial system, then, you need data access that works globally. That's something that would make me very proud—if we can enable that, even if it's for a small percentage of the population that actually have global lifestyles and assets.

It's really just in an early stage startup phase of making sure we build a product that has 10 customers—that's our next milestone—10 companies rely upon us for their infrastructure, and then launching it to get broader access.

We think we're building this obviously in the open, publicly, by the very nature of open source. We are bullish of the open source community being a unique acquisition channel for us, a developer-first open source community.

Disclaimers

This transcript is for information purposes only and does not constitute advice of any type or trade recommendation and



should not form the basis of any investment decision. Sacra accepts no liability for the transcript or for any errors, omissions or inaccuracies in respect of it. The views of the experts expressed in the transcript are those of the experts and they are not endorsed by, nor do they represent the opinion of Sacra. Sacra reserves all copyright, intellectual property rights in the transcript. Any modification, copying, displaying, distributing, transmitting, publishing, licensing, creating derivative works from, or selling any transcript is strictly prohibited.