

Rebecca Allen: [00:00:00] The initial design of the typo tool, it was sensitive to that, right? So it would say, if it is, we know this is wrong, then change it to, we know this is right. But, um, over time it became clear that even in those cases where they did fix it, sometimes the fix was not really a fix, but then our, our change would go away because it was sensitive to what it was at the beginning.

So we eventually just was like, no, no, no, no, no. If we fix it, we know what we wanted. And so that's what we do.

Dave Schappell: [00:00:37] This episode is sponsored by skill jar.com founded by ex Amazonian, Sandy Lynn and Jason Stewart skill jar is the leading customer education platform, innovative companies, such as Tableau, LinkedIn, and hundreds, more rely on skill jar to onboard, engage and retain their customers at scale to learn more, visit skill jar.

Dot com S K I L L J a r.com. Hi, I'm Dave Chappelle and I'd like to welcome you to the invent, like an owner podcast, where I talk with the Amazonians who help build amazon.com into one of the world's most valuable companies. This weekly podcast is for entrepreneurs, business leaders, and all students of history.

The goal of the podcast is to capture the Amazon creation stories and create a historical archive. On that note, my guests are recalling history as best they can. It's possible. Some of the details are fuzzy or just plain wrong. And if that happens, it isn't intentional. And I invite future guests or commenters on the site to help us get the facts as straight as they can be now on with the show today, I'm excited to be talking with Rebecca Allen who started at Amazon in March, 1996 as a software engineer.

And is Amazon famous for being the creator of the ASEN and a we'll tell you about what that is today and why it's really important, uh, for Amazon. Many people don't think about catalogs or website catalogs, but Amazon's catalog is one of its most valuable assets. And over time enabled it to outmaneuver companies like eBay.

And Rebecca's going to share some of the details that went into creating Amazon's catalog. And she's also going to share some fun facts about how many unique Asians have ever been created and how that is accelerating rapidly. Uh, welcome Rebecca.

Rebecca Allen: [00:02:23] Hello?

Dave Schappell: [00:02:24] So before we get into the catalog and ACEs and all that, maybe just tell the story. Cause it was you were, you were there so early, uh, tell the story about how you got to Amazon, like who, who the people were that you sort of joined with and. You know, sort of leading into what were you hired to do?

Rebecca Allen: [00:02:41] So my first relationship with Amazon was as a customer, not as an employee. So in December of 1995, I was trying to find a book. I believe it was confessions of a poker player to buy as a Christmas present. And, um, I have a lot of trouble getting it ordered through university books, which is an excellent independent bookstore in Seattle to this day.

And, um, they tried really hard, but not all small publishers answer their phone consistently. And so after that failed, I looked around online because that's kind of what I did back then. And I discovered Amazon and I was like, what the heck? I'll try this. And I ordered it and it gave me a very, very pessimistic timeframe on when it would arrive, but it never, I'm very, very promptly.

I was excited. I gave it to the person that it was for. And, you know, didn't really think a lot of it after that. But then a few months later that person actually was interviewing at Amazon and Scott. Talk to me about it. And he was super excited about how smart everybody was there, how much fun they were to talk to.

And I was like, wow, how would you feel if we were both working there? And he was like, that'd be great. So, uh, I couldn't work with the same recruiter who had recruited him because I had used that recruiter on a previous job, 11 minutes earlier. And so I thought, Hmm, well, I gotta get my resume to them somehow.

And I figured I'll just send it to them in the HTML because it's a web thing. So maybe they'll like that. And so I got an email back from Nicholas Love, joy, and I came in for an interview and they wanted a code sample. And I said, well, most of my code is owned by the companies that I have written it for.

However, I have this thing that I wrote for my own website, which is age book catalog. So you can take a look at that. They like that they. Shell thought it was pretty humorous. That that was the code sample I gave. And he wondered if I had written it for this job interview. And I was like, no.

Dave Schappell: [00:04:41] It was just the books you owned. Is that what it was the catalog.

Rebecca Allen: [00:04:44] So it was a few thousand books in it and that's all.

Dave Schappell: [00:04:46] Yeah. So that's great.

Rebecca Allen: [00:04:49] So, uh, I got hired in, um, also another coworker at the, my previous job I was at when I was interviewing, um, another coworker was also looking for work because that startup had been bought by a larger company and it didn't really feel startup B anymore.

So, um, Ellen was also looking for work. And so Scott, Ellen and I all started at the same time, which was really kind of nice because we knew each other to one degree or another and we got along well.

Dave Schappell: [00:05:18] That's awesome. And were you hired to do, what were you hired to do? Was it to work on catalog or was a bunch of things or, you know.

Rebecca Allen: [00:05:25] At the time that we came in, there were two major programmers, um, shell, captain, uh, who would be there. Even after I left, um, and pelvis Davis, his name was Ben Barton Davis. It is not anymore. Um, he, uh, was leaving, um, Chell had previously done. They had kind of split stuff up each kind of doing half. And, uh, shell had more focused on kind of the front end tools and pover data's on sort of the fulfillment end of things.

Yeah. Um, and so Paul was exiting and so the three of us were going to be basically replacing Paul plus because obviously growth was happening very quickly. Right.

Dave Schappell: [00:06:08] I'm just gonna throw in real quick for listeners. There's a really good interview with shell on the internet history podcast. And he talks, if you're interested in going into that detail, he actually talks a lot about the, which parts he focused on and which parts Paul focused on.

And I'll put a link for that in the show notes, because it's, it is really well done.

Rebecca Allen: [00:06:26] That's awesome. So, um, when the three of us came in, I was pointed at catalog and search, although I never did anything with search. Um, Scott went and did credit card processing and Ellen, um, worked on the extensive work that needed to be done with the backend tools to support the rapid increase in.

And I mean, cause that was a very labor, intensive minute. Very things could go weirdly wrong with fulfillment. So that was.

Dave Schappell: [00:06:57] She, she handled warehouse software and all that sort of stuff. Yep. Okay. And so, so we're at catalog, let's just, that's where we're going to spend a lot of the time today. So talk about, uh, how Amazon, like in, in language for that, my mom can understand, talk about what, what, how the catalog was built and why it was difficult to do, you know, uh, and, and as it got more difficult over time, as we added product lines, that sort of thing.

Rebecca Allen: [00:07:25] So, um, we, you know, we wanted to sell all the books and, but we also wanted to sell all of the books. So our catalog was a combination of catalogs of books that we could sell, which sounds a little bit circular, but basically we ordered books from distributors initially. And so we had a catalog of what they could sell

us, just like any restaurant buys from a restaurant supply catalog or any, any other business.

So we had, but those catalogs are not customer facing. They need to be good enough to order them. They don't need to look pretty. And so there's a lot of errors in them. Sometimes they have working titles rather than the actual title and they don't always spell authors names correctly. Right. So what we did was we combined based on the ISB, which was the standard numbering system for books at the time.

Um, we combined them and we produced a customer facing catalog, but when I arrived, there was really no way to fix those errors.

Dave Schappell: [00:08:27] So we would get the catalogs from, was it baker and Taylor in Ingram or was it somebody else? They were the two big ones. And so we would get the catalogs from them. They would ha they would be, it's kind of like the, what I would think of it as you walk into a warehouse where you're going to buy stuff in bulk, and it's not quite as pretty as an apple store, you know?

And so we'd get those, those CDs or where they, well, they had to be CDs. We weren't typing the stuff in. Yeah. And then.

Rebecca Allen: [00:08:51] And they were not delivered over the internet yet. Yeah. Later that would happen. But yeah.

Dave Schappell: [00:08:56] So we would take them and we would merge them together.

Rebecca Allen: [00:09:00] Yeah.

Dave Schappell: [00:09:01] Yeah. And then, so what would we do when there were mistakes? Do we call them and have them fix them? And they promptly sent you a CD.

Rebecca Allen: [00:09:09] They did like us a lot because we ordered many, many things from them and we didn't return anything, which made them very heavy from a business perspective, but that catalog was good enough for ordering and that was all they cared about.

So I'm not a hundred percent certain how much we really wanted to solve things in that kind of way. Anyway. So when I arrived, there were 10, there was no way to fix anything. I mean, there was, but it was very cluggy and it was labor intensive and potentially dangerous. So, um, what, my first one of my early jobs was to create a tool that somebody like a customer service person, but it was a separate department over time and called catalog.

Um, could call up an ISBN they could pull the ISBN off of the web. They could enter it into the tool that would give a very, very rudimentary display. And it would give you the option to change on a field basis, like the author or the title or whatever. Um, initially it could be sensitive to what was in it.

So only change it based on the prior value, but over time, all anybody used was to put the correct value in.

Dave Schappell: [00:10:09] Right. And so, did we get most of these requests or to fix things from customers or was it from the authors, like who, who would be making these requests? I'm assuming they just got more and more frequent as we got bigger and bigger.

Rebecca Allen: [00:10:22] Very, very true. Um, a lot of authors care, a great deal about how their books are presented to the world. They go to bookstores, not just to do readings, but to make sure their books are there on the shelf and displayed attractively. And so what something looks like on the Amazon detail page is the rough equivalent of what it looks like on the shelf.

So they, they care a lot and some of them are quite entrepreneurial and they would call back every few days until we got it fixed, which where there was a way to fix was a problem.

Dave Schappell: [00:10:50] I mean, we all care about our name being spelled correctly. Absolutely no type of like it is funny. Cause when you nowadays there's so much, there's so many products that get created so quickly on Amazon, that when you do see a bunch of typos, you're like, eh, maybe this isn't the real item.

You know, this might, this one might be a counterfeit that I should, uh, go take another look at. So that, that makes sense. And so you would have those two. Catalogs. And they would come every week or every month or whatever, and then you'd have your own database of fixes, but how did they not overwrite one?

You know what I mean? Like how did you make sure your fixes stayed constant? And like what happened then if they did make a fix in the catalog that they were sending you turf, it's getting a little circular.

Rebecca Allen: [00:11:30] But like was sort of the conundrum at the beginning was, well, do we want to use their fix if they do fix it, or do we want to make our fix the definitive fix the initial design of the typo tool?

It was sensitive to that, right? So it would say if it is, we know this is wrong, then change it to, we know this is right. But, um, over time it became clear that even in those cases

where they did. Fix it. Sometimes the fix was not really a fix and, but then our, our change would go away because it was sensitive to what it was at the beginning.

So we eventually just, it was like, no, no, no, no, no. If we fix it, we know what we wanted. And so that's what we do when the catalog was built. Um, it was essentially walking through the ISBNs and the order on the databases, um, and merging sort of what we had from whatever catalogs that we were using distributors, but later publisher information, and then producing a record to go into our catalog.

We didn't rebuild our catalog from scratch. We built it incrementally.

Dave Schappell: [00:12:37] Yeah. And so, and, and then, so basically this merged catalog that you came up with. Right. You know, repeatedly that's what, uh, the, that we use for search and browse and everything else. Uh, correct.

Rebecca Allen: [00:12:50] And all the warehouse tools that looked anything up by eight ESPN. Yeah. Everything touched it. Yeah.

Dave Schappell: [00:12:56] Got it. What was, uh, just as a fun fact, do you remember what the name of that database? What did we, what was that named by any chance?

Rebecca Allen: [00:13:04] I think it was biblio dot DB.

Dave Schappell: [00:13:06] Yeah. That's brings a bell. That's why I was wondering. Um, okay. So we have, we have the catalog now. Let's talk about and we have books.

So a big part of this story is, or one of several here is the creation of the ASEN and you're like, what's the essence. So you mentioned that we were using ESPN, which I think stands for an international standard book number. Oh, actually I had a question about that. Those are issued by a company named Belker and I think in the United States, I think we should breast, we should settle it for the record because there's rumors that we bought bulkers database at some point, or now we're.

Rebecca Allen: [00:13:43] We didn't, we didn't buy books in print ever.

Dave Schappell: [00:13:46] Yeah. We never paid for books in print. We basically built the catalog ourselves and improved it over time and that's never changed. And, and as far as we know, it's never been licensed to anybody else. So it's Amazons. So, so talk about why we needed a replacement for like maybe why ISBN was chosen to begin with, uh, as right or wrong, and then why it needed to be replaced with something else that, and, and what we ended up with.

Rebecca Allen: [00:14:13] So I'm going to actually go back a little bit, even further in time and talk about what is, you know, kind of, how did the ISB come into existence? I haven't researched this recently. I'm going off of memory and may get it slightly wrong, but the ISBM was not originally an ISB and it was originally an SPN, which means that it wasn't an internationalized.

At first. It was something that, to the best of my knowledge was invented by Waterstones. They needed a way to, you know, have a unique number for the books they sold.

Dave Schappell: [00:14:41] Waterstone is a big bookseller in the United, in the UK, similar to Barnes and noble in the U S yes, sir.

Rebecca Allen: [00:14:45] This was essentially their solution to the same problem, but a long time ago, it got taken off, uh, over by standard spotty and, and internationalized.

And so within the U S Barker is the one that controls issuing the ISBNs.

Dave Schappell: [00:15:00] Got it.

Rebecca Allen: [00:15:01] So if you're building a book catalog of like all of the books, it is quite natural and, and fitting in with people to sell them, to catalog them by ESPN. Um, it gives you a unique identifier, at least in principle, and it's an identifier that's already associated with the book and can be compared to other sources of information.

What about books? Great. Unfortunately, charges what small publishers in the United States regard as quite a lot of money, or at least they used to, to get an ISB in. So they didn't necessarily you'd buy range and they didn't necessarily either understand how an ISBM was supposed to be used. Um, or if they did, they didn't necessarily follow the rules.

Right. So, um, it's supposed to be basically a night, not attended at number, but a nine digit number with a check sum at the end. And, um, sometimes they would just number through the entire range they had, because you could get 10 times as many, or yeah, you can get 10 times as many numbers for the same amount of money.

Also, you know, reuse, recycle, they would reuse it. I as pans, which is very unfortunate because if you sold somebody, you know, particular ESPN, and it was like how to build the shed, and then you sold the same ISP into somebody else. And it was like how to build a chicken coop. There is going to be some confusion.

So there is a lack of unique neatness and yeah.

Dave Schappell: [00:16:29] Confusing to me. Cause you would think part of bulkers job would be to be the, uh, the complete list of ISBNs with related products and that they would make sure there weren't two products with the same ISBN, but we don't have to go into that rabbit hole.

We suffice it to say it was, it was not a perfect solution.

Rebecca Allen: [00:16:48] So, but, and, you know, and she'll realize that pretty much immediately when I first interviewed this was a problem that he brought to our attention. So it's not like.

Dave Schappell: [00:16:57] Well listener for the listener shell was Amazon's first seat, uh, chief technology officer. So he was sort of employee number one, we after Jeff Bezos. So, uh, shell was very involved in all this. So you'll keep, she'll keep, Rebecca is going to keep saying shell, but shell is the CTO. And then which is, yeah, go ahead.

Rebecca Allen: [00:17:14] So he knew this was a problem, but what are you going to do? So it became a really pressing problem when we wanted to sell products other than bugs.

So the motivation for ACEs was to become, not just all the books, but all to everything. And when we did that, there were some ideas about how to list items in the catalog that were not books. One idea was, well, we'll just make up our own a ISBNs or possibly buy ISBNs from. Belker, it would have been cost prohibitive to buy the number of ISBNs that we were ultimately going to want to use.

And also they would not fit into the namespace of those nine digits plus the checks. So, um, so we needed something else, you know, there were my proposal. Should I get into my proposal at this?

Dave Schappell: [00:18:07] Yeah, this is where we're getting nascence.

Rebecca Allen: [00:18:10] So if you want to back, uh, if, so the core issue is that our entire collection of software, the catalog and everything that touched the catalog expected the key for an item in the catalog to be a 10-character string.

So that was all over the place in the catalog, but it was also all over the place in every single order. And if you're going to change a whole lot of code all at once, You would like that change to be as controlled as possible. So as to minimize unexpected consequences, So we really, didn't want to change it from being a 10-character string, but we knew we could not accept the limitations of an ASIN And I looked at that and I said, okay, well, if you want to pack something into a space like that, your best bet is to use a larger base. So if you think about base to each space in a number, it can have

two possibilities, zero run based 10. In that space, you can have nine possibilities in. You can, if just using the alphabet and upper and lowercase, plus the numbers, you can have a really big space and have 62 possibilities in each place.

Nobody liked this idea because they didn't like the idea that. The number B capital B followed by zeros all zeros was going to be a different ASEN. Then lowercase B followed by all zeros. They, that was an unpopular the solution. So then I said, okay, we could do base 36, which has all the letters of the alphabet case insensitive.

And the digits zero through nine, people were fine with that. And we had some discussions about, okay, what is good or bad about this proposal? There were some people who wanted there to be some structure to the identifier. So that, for example, maybe you could code what type of product it was in one of those spaces.

And I resisted that pressure because I felt like that would limit the ultimate amount of products we could fit into our numbers system. Right. I won. I figured if we're going to sell it, you know, lots of things, then we, for a long time, hopefully then I don't want us to run up against this limit again.

Yeah. So, so, um, because people thought they might want special Asians for something I said, all right, fine. How about you have everything that starts with the letter a and then we'll just start numbering this at B followed by all zeros and we'll just run a counter. They won't mean anything at all, but illegal ESPN starts with B followed by all zeros, and then it goes to the same thing, but one, and then two, and then et cetera, counting upwards to ever greater numbers.

Dave Schappell: [00:21:07] It made me just go type in amazon.com/b with nine zeros. But that got me to a, uh, nothing, nothing there. So just figured I figured I'd check. I was interested. And so, uh, that's, that's the, that's basically the Genesis of what became the Amazon standard identification number or an ASEN. Um, can you tell about like, was that quickly that name quickly, uh, Quickly adopted and enjoyed, or did people need convincing?

Rebecca Allen: [00:21:34] So the whole project needed convincing. This is going to be an enormous change and it was, we really didn't want to have to do this change twice. And it needed to be very, very, very, very flexible because we couldn't anticipate all of the products that it would ultimately need to support. So, um, in addition to reserving the AA space for special purposes, people were worried about backwards compatibility.

And so I also said, okay, if it's a nice BN, that is an ASEN already, they got grandfathered in. So the next step was, um, the CTO shell had to, if this was going

nowhere without his approval. So, uh, I felt humor would help bridge a difficult gap of, you know, changes, bad changing all the code all at once is worse.

And. Honestly, no, I don't think anybody else in the planet has an item number in a base. So it's all a little suspicious. Um, I felt like the fact that we had to cram all this into 10 characters when ordinarily, if you need more space for your skew or your item number, you just have more space. I mean, you just have more digits.

Right? Right. Um, so I felt like it was a sin that we had to do this at all. And that was part of how I sold this to people is, look, it's a sin that we have to fit into this space, but we do. And here's my solution. But for shell, I felt like we would need something a little bit more. I don't know if technical is the right word, but we needed something maybe a bit more Mathew.

And so I told shell that we would pronounce it arc sine.

Dave Schappell: [00:23:16] So an arch sign being a mathematical term that appealed to appeal to shell.

Rebecca Allen: [00:23:20] I thought that was humorous. And that got us started on the conversation, which was the most important part.

Dave Schappell: [00:23:26] And so based on. Based on that. So by the way, for listeners, if you have never done this and we're having never noticed it, you can go to almost any product on Amazon.

I'm a, of sometimes accept the book because the book has ISB N but you can just go to the page troll F and S uh, type ASI, N uh, or look in the URL. And you'll, you'll often see it there. It's sort of a unique identifier and we'll get to it in a bit, but one product page on Amazon can have multiple variations.

And we'll talk about how so one product can actually have multiple agents. And, uh, I may be misusing terms there, but it's just a, another complication of things that we're going to talk about here in a bit. So, can you, do you have any idea? I mean, based on, based on, uh, based 36, can you tell, well, how many items there are in the catalog ever, or how many ACEs have ever been created.

Rebecca Allen: [00:24:15] Yeah. So, you know, if you think about it, if it's just the serial number and a recent serial number, you can kind of guess how many serial numbers have ever been used. Maybe they weren't all shipping products. Maybe they didn't ship for very long. Maybe they were one-offs of some kind, like a used item that was only one physical thing or whatever, but, serial number is a serial number.

It's numbers counting up. So if you take a recent ASIN and you take that and put it into a base 36 to 10 converter, and you take that number out, and then you subtract B followed by nine zeros, run that through the base converter, take that, subtract it from the first number. You will get a sense of the total number of ASINs that they have gone through.

And it's a really big number. I do this every few years. When I did this in 2016 and left notes on my blog, it was, I dunno, like 111 billion, which I thought was a pretty big number. And now it's like over 700 billion and I obviously had to lie down for a little while because I was like, what are they doing over there? Sorry.

Dave Schappell: [00:25:18] Do you think the number is accelerating? Like, I mean a hundred billion already seemed high.

Rebecca Allen: [00:25:23] Yes. Those are like 600 billion more in like five years. Like what? I mean, that's speeding up. So, yeah, I don't know for sure. I don't work there. I haven't talked to anybody about this, so this is not insight information at all, and I could be completely wrong.

I am a, an armchair speculator at this point. Um, my guess is it's either, it's probably expiration dates because if you want to have an expiration date on a product and believe me, if you're selling groceries or. Possibly even batches of medicine, you're going to need to have an expiration date associated with the product.

Um, this may matter for recall purposes, uh, but also just, you know, inventory management, if you're doing fulfilled by Amazon and you want to make sure that stuff gets pulled at the right time before it's gone bad or whatever. Yeah. So it, I know that they support expiration dates. I went and read some of the vendor pages at Amazon.

And if you have an expiration date, there has, can only be one expiration date on an ASIN. So that basically means that like anything, like say lettuce, um, that has an expiration date. Um, you're gonna like need a new one every little bit.

Dave Schappell: [00:26:29] And for the listener, um, uh, Rebecca was nice enough to do two blog posts that we're going to put up coincident with launching this episode.

And she goes into a lot of the detail of this. She actually went and pulled, uh, Brad Stone's new book, uh, Amazon Unbound, which is being promoted everywhere right now. And estimated it's about the 700 billion, uh ASIN uh, you know, in that range. But she also highlighted that even the difference between like his Kindle book, his hard cover book, his audible book, even those are off by millions, um, because of products that are, you can.

Rebecca Allen: [00:27:03] Billions or a billion, 4 billion between the Kindle and the audio book, unless I did my math wrong, which I feel confident, one of your awesome listeners will bring to my attention if so.

Dave Schappell: [00:27:14] All right. Well, so, so that's the ASEN and you can, uh, you know, for the listener, you can see the complex, the complex thought that went into thinking about this, and the fact that something that was created back in 97 is still being used today and hasn't run out of space means it was pretty good. Uh, pretty good forethought.

One, one question before we move on to the next big topic, we're we, the only ones wrestling, maybe we were because Barnes and noble had their own in-house solution, but like w do you, did we emulate anybody else? Or was this all blue sky thinking, you know, for a problem, we knew that we were going to have back then?

Rebecca Allen: [00:27:53] As far as I know, I don't know of anybody else who does a different base to wedge an item, number skew, or catalog number into a constraint space.

I know of nobody who has ever done that. It's a very odd solution that came about because. Um, we had a historical constraint for us. We've got a code base that's highly dependent on the 10 characters of the eyes being, and we needed it to be a lot bigger. Right. So, you know, everybody has to have a skew or item numbers or catalog numbers.

Um, you know, we had, we had more earlier than like anybody, as far as I know, Amazon had more distinct item numbers than anybody earlier and maybe ever like probably, I don't know.

Dave Schappell: [00:28:44] Well, so let's move on to the next, I mean, it is, it's just the thing that I found most interesting in this topic. Cause I knew what acents were, but I didn't know anything about base 36 is just like.

Rebecca Allen: [00:28:56] It's not in the Wikipedia page. I want it in the Wikipedia page.

Dave Schappell: [00:29:02] Yeah. Awesome. Well, yes. And I was just thinking like the fact that it has. Sustained. And it's still in the bees that hasn't moved to the seas or the use of these means that, you know, it was pretty, pretty good. Yeah. Uh, yeah, but the numbers are increasing by billions by the day.

So, you know, it'll, uh, it's a race against time. So let's talk about the next big topic that you brought up sort of on our way through these catalog problems is biblio record. Can you explain what that was and why it was why it was a, a big challenge?

Rebecca Allen: [00:29:34] So one of the last things that I did when I was at Amazon was I changed the fundamental data structure, both in the code and in the database.

Um, as you can imagine, the initial data structure that we used in the code and in the database was very book centric. So there was a spot for a title. It had a certain length, uh, that was pretty easily changeable, but there was a title in the database. There wasn't space for multiple titles. There was a space for.

An author and then later multiple authors and eventually, you know, we added things like, you know, how, how much did it weigh or how many pages did it have? But it was all very book centric. So if you were like thinking like, okay, we're going to sell CDs, we're going to need probably track listings or something in there.

Um, we're going to sell maybe clothing someday. You're going to need to have color and size. And who knows. Yeah.

Dave Schappell: [00:30:29] And for the, for the listener, like you just mentioned, wait, but I'm pretty sure dimensions are in there as well. Uh, and the reason obviously is, so that way the system knows what size box does this thing have to go in, or how heavy is it might need a different type of box or different type of shipping, you know?

So all that type of stuff was being thought about how do we get that stored in, in the catalog?

Rebecca Allen: [00:30:50] So, um, the, the initial design of the database was a very, very simple database. It wasn't like a relational database that you might be thinking of that lets you search by any number of fields. Um, It was a key value database.

The key was the ISBN later ASIN and the value is just a series of tabs, separated, new lines terminated is that so, um, if you wanted to put any type of field in there, you both had to change the code to know what to do with it, but you also had to change the database so that it would be in the correct location in there, because it was expecting these things to be boom, boom, boom, boom, boom, all in a row.

And I wanted to make it so that there was some more flexibility. And I looked at a variety of, um, cataloging schemes, including, uh, the library of Congress is Marc records. And I really liked their idea of, instead of just having individual fields been, been moving them pairs of fields where the first part of the pair was a code to tell you what type of field two weeks.

Expect, and then the value that was there. And that was the basis for a redesign that let the Biblia record store kind of anything and add more with less dependency on a brand new database being rolled out simultaneously with the new code that sort of let you basically roll them out without synchronizing that, which would make operationally rolling out a new catalog, much less fraught.

Dave Schappell: [00:32:24] So we would basically see that code and it would probably check a separate database or a separate team, separate table, whichever table of all the codes and what to expect. And that way you could just add codes, if you want over time, as you, as you discovered they're needed. And then the main, uh, maybe using terms wrong here, but that main database, or that could just grow and grow and grow with codes and, and respect to values and sort of, uh, it didn't need to be rebuilt every time or redesigned.

Is that kind of basic understanding?

Rebecca Allen: [00:32:54] Yeah. Now, obviously, if you didn't have any logic in the code base to cope with a particular type of field, right, you had to add that logic for each type of field, but this would at least mean that the routines that read in the database didn't have to be changed as often and would not break unpredictably.

If you rolled out a database with fields that it didn't know about, it could just ignore them in a clean fashion. Got it.

Dave Schappell: [00:33:24] And so was that, that did that problem surface, just because you were frustrated with new requests that were coming in for things to put in the catalog, or was that something that you just foresaw was going to be a problem based on thinking about the catalog all day, every day, you know, for, for a couple of years.

Rebecca Allen: [00:33:44] The primary motivation was new products.

Yes. However, this is also essentially a refactoring in software engineering practices kind of project, right? So if you like have a system that every single time you want to make the least little change, you'd have to change two really big things and update them simultaneously. That's scary and bad, and there's a lot of potential for bad things to happen.

And so you really want to replace every time you're doing something like that. You want to replace it with something that lets you separate those hazards and that lets you, um, you know, test them independently that lets things like. If you get into an unpredictable state as to which one knows about which that they still behave well.

Dave Schappell: [00:34:28] Yeah. Perfect. But it does help. So, so then when we, the next big, the next big catalog and I'll call it development, you can better title it is his title authority. Can you explain what, why title authority was like, was needed? Did you, and, uh, maybe describe what it is first and then, because I have a lot of questions about title authority.

Like when I went back to Amazon in 2012, we were still referring to title authority. It was in a different context. So I don't know if they still use it for the product catalog anymore, but, um, it might not be a term that a lot of other people know about.

Rebecca Allen: [00:35:03] Uh, so like that was one of the last things I did.

And I honestly have no idea if they're still using it or if they devise something different, but I can tell you what I designed, what happened with it may have used what I designed. Um, I, you know, you probably asked him cone. He would probably know. Um, he was there for a long time after I left. So he would know whether that got used on what it got used for.

Um, title authority comes again from the library of Congress. basically the library of Congress is a lot of people, so librarians are catalogers. That's what they do. And, they don't like it if you have John Smith and Jonathan Smith. And you're not sure if John and Jonathan are the same guy or different guys, they would kind of like to know, well, which one is it?

And we want to sort of have all of this person's variant representations of their name points to the definitive representation of their name. So that's kind of the library of Congress thought process on it. I actually thought about title authority in a slightly different way. I was not as concerned with, the definitive representation of the name, but I was very concerned with ensuring that the catalog could capture and.

Help the customer with relationships between products. So for example, if you want to read all of the Harry Potter books, you would like them to somehow be connected to each other. And one way that you can represent that is by representing each Harry Potter book as being part of a series. Yeah. And, and so, so anything that lets you like say, okay, um, Janeane Krentz is the same person as Amanda quick.

Right? So to use an author who writes on your multiple pseudonyms, so you'd like to be able to help the reader find that writing it or anything.

Dave Schappell: [00:36:59] Yeah, go ahead. I'm sorry. You've got, I just like, I think this is really fascinating. It's not something you really think about. I mean, it's easy to think

about Moby Dick is a book and there's a hardcover and a paperback and an audio book, and those are, what do we call that?

It's like same product, same work sort of thing. And. Just to put this in context, you all had to create all of this, like this didn't exist anywhere in, uh, in, or did it exist at all in some of these, these catalogs we were getting.

Rebecca Allen: [00:37:27] So, um, at least one of the distributor catalogs had a related ISB in field.

Um, which I don't think, I don't know if he ever displayed that information and I'm not sure how much we ever used it. I may, I'll probably wake up in the middle of the night going like, ah, I remember that. Yeah, but I don't remember right now. So, um, but so in terms of like even series information, um, and, and pseudonym information absolutely was not.

Present anywhere that was just kind of stuff that I honestly, sometimes the Outfitters would just as soon you didn't figure that out.

Dave Schappell: [00:38:05] So well, other examples you gave me was, uh, I may mispronounce it, but like Ilona or, and answers yet. You're like, oh no.

Rebecca Allen: [00:38:14] It's a husband and wife writing team. Yeah.

Dave Schappell: [00:38:17] It's a husband and wife riding team under, under a different name. And you all had to figure that out because somebody might search on the husband's name or the wife's name or.

Rebecca Allen: [00:38:26] So that's not a particular problem there, but I always feel like, um, if you can provide a mechanism for, for helping. So a lot of readers of books, they really want more by the same. Yeah, whatever.

And so if they've read everything by one author, but you can show them that, oh, actually this person is writing under this other name or they write by themselves here, but they write as part of a writing team there, or they write by themselves now. But they used to write as part of a fictional name that multiple people wrote under like Carolyn Keen.

So, so anything you can do like that to help the dedicated reader customer find more of what they want. You know, this was before we got really awesome with recommendations. So, you know, this is what.

Dave Schappell: [00:39:17] The reason that this is a really interesting is you can just see how much you and you're like, You know, probably millions of people, hours working on these problems and getting these associations and getting the ACEs and getting the other tools we were just talking about, like, just getting this all to work is what makes sort of shopping and browsing and exploring and discovering it wouldn't work.

If somebody wasn't sitting there getting all those details in those relationships, right? Like the recommendation engine wouldn't work as well. If it didn't have all this information figured out.

Rebecca Allen: [00:39:52] Well, so I dunno, I don't know if the recommendations engine works in a very different way. And so I, but, but what I guess I would say is that, um, The unsung heroes here are, it's not me.

I was only there for a very brief period of time. Right. I, I put some errors into this. I thought hard about it. I talked to people about it, but it was, it was pretty compressed, right? When I did it, as I created a framework for capturing these relationships, that was extensible. Well, actually, by the way, but then had to happen is, is so many people had to populate that information and the value to Amazon of that information being captured and then available to the customer is phenomenal.

Dave Schappell: [00:40:29] Yeah. And I was actually saying, I was actually referring to the entire catalog organization. Like it's this massive team that there's people in every country on every, you know, in every store, like, and it's not something you normally think about. I think most people think about customer service and the people in the warehouse and the engineers, but like, there's the amount of detail that went into getting the catalog.

Right. And all I meant about personalization by the way is just if person searches for the partner, we probably have. 119 accents for the, actually the partner, you know, like all the different variations, but they know which like main one to put up there somehow and again, without getting into it, it's it, it just shows how complicated this stuff gets.

As you start, you know, you start thinking about the problem of displaying a single piece of work, you know, and then you're like, well, do they want the audio book? Do they want the album? Do they want the hardcover paperback? Like, it's really complicated. So, um, so T uh, so all of this, we, you know, we now have billions of these ACEs that have been created and cleaned up and edited.

Uh, has to your mind, is Amazon ever licensed this ASEN catalog out or is it all it's been built and it's primarily used on Amazon?

Rebecca Allen: [00:41:44] Again, I was there briefly. And so a long time has gone past that. I know absolutely nothing about. What might have been done with that? I'm certainly never seen Asians anywhere outside of Amazon, except there was a period of time when there was a deal between target and Amazon for right.

Um, yeah. And so Asians were appearing on the target store for some period of time, but honestly I stopped paying attention a really long time ago, so I have no idea what's going on with that. Got it. So, uh, Well, I did an earlier episode. I don't know if you listened to the Jane Slade, Colleen Biram episode.

I have not yet. I am looking forward to it though.

Dave Schappell: [00:42:23] Well, there was lots of it. Yeah, I know. And there was something that came up in it that I wanted to ask you about, because I think it relates to catalog and it talked about, um, Jeff introducing at some point some massive amount of books that were like, not in print anymore, or, you know, he just said, Hey, I got this catalog of a million books that we know were printed, but you know, we're not sure available.

Does that, did that happen?

Rebecca Allen: [00:42:49] Yeah. So that was the original rollout of used books. And, um, you know, I think, I think one of the things that's kind of awesome about early Amazon is how much of it was just smart people making it work. So, um, Jeff got the library of Congress database or some component of it.

And he was like, let's just list all these books. We know they're real books. We'll list them as youth books and say, you know, hard to find for the availability. And then we'll just basically, if somebody actually places an order for it, we'll just, you know, you know, open up the phone book and like call all of the used bookstores and see if we can get it from them and then sell it.

Right. Which is like bizarrely, but also like totally insane. Yeah. Right. It's inspired by the fact that if you look at, um, what a used bookstore pays for product, you know, back in the day, there used to be used bookstores, you would go to in person, in your local university district and you, they would like buy your book for like a quarter of the cover price and they would sell it for half of the cover price and, or some.

Variation on that formula. Well, what do you think about it? That is a pretty awesome margin and at least it looks really awesome until you realize what exactly is involved with sort of managing this whole process. And then it stops looking like an awesome

margin at all. And it just looks like a lot of work, but you know what, that's all in the future for this story.

At the time we had the library of Congress database and we had a bunch of people who could make phone calls and Jeff is like, let's do it. So I, you know, I took a leave of the database and I said, okay, so number one, a lot of these things don't have ISBNs and, but some of them do, and he's like, okay, we'll start with just listing the ones with the ISBNs.

How many is that? And I came back and I said, it's this number? I don't remember what it was, but it was a big number. And so he says, cool, because, you know, he was just trying to get the catalog really big. He wanted to, you know? Yeah. So, um, I was like, okay, so I can get this in. They're going to look a little weird and we're going to have some Y2K problems, but we need to fix those anyway.

So, okay. And I said, I know how long it's going to take me to do this. This is a very well-defined problem. Right? I need to take this database and I've looked at it and I need to bash it into something. That'll go into our database. It will take this long. I can roll it out in this timeframe. However, if you do this to the rest of our system, it's not going to be as predictable.

And so, you know, you're going to have to ask somebody else about Alonda, it's going to take. And they were all like, oh, it won't be that hard. It'll be like this, this and this. And I'm like, oh no, I don't know exactly. What's going to blow up on your side, but something bad is going to happen over there. And, and there was some disagreement about this and I was right and they were wrong and I held it over their head forever.

Dave Schappell: [00:45:31] So, Rebecca, uh, you warned everybody that things were going to break, uh, with this introduction of, of used books. You weren't sure exactly what was going to break, but you thought things would break, uh, did something break.

Rebecca Allen: [00:45:44] Yes. And I really, honestly, I might be the wrong person to get the details from, because I only know what broke at a very high level, but I do know that things broke both in, not my database in the database where order information lived.

And also things broke with, um, process throughout customer service and, and elsewhere. Obviously whenever the order database breaks down that impacts everything, right. It also impacts the customer because it makes it slow to place an order. Um, my piece was very controlled, right? It was a matter of taking one batch of information and then taking, extracting the parts we could use and getting that into the catalog.

That was something that I could assess ahead of time and control. However, the rest of the order process involves humans and also, um, it involved, uh, really. It involved a lot of tables in a relational database, which had been accustomed to a particular type of order flow. Right. Normally when people placed orders on the website, um, even if it was a special order boat, like the one that was my very first order at Amazon.

And even if the expectations on fulfillment were not particularly optimistic, you are going to get that book in a few weeks. Right. Almost always. So that table would clear quickly or relatively clear. We didn't, we didn't do pre-orders back then. So you didn't have like millions of people. Pre-ordering the next book by their favorite author.

And then having that order sit in a table for months or a year until the book actually was published and could be delivered, moving through the process of tables and then finally sort of sitting somewhere to be accessed by order history. So basically there's tables blew up and everything's slowed down.

And every single tool that touched that relational database ran really slow. Right. And until that database was tuned to better cope with the change in size of tables, and then quite rapidly, shell rolled out shell and Ellen rolled out a solution that involved, um, essentially splitting up many of those tables so that the orders which were sitting right, would be in a table that was not commonly accessed.

Dave Schappell: [00:48:00] And I wondered if they would split out those, even the rare book orders to send these are gonna, uh, you know, work differently. So, but, but the nice thing is you got to walk around for the next year or two saying, I told you so. Right.

Rebecca Allen: [00:48:13] And I, I use that power for good. I swear. Um, I mean, there was a little bit of, I told you so, but the, I told you, so it was mostly directed at let's think before we roll out a big new thing, right.

Let's not think about, okay, well, this is obvious. So that must be all of it. But let's kind of like try to walk through what is going to happen in the process, and then try to imagine how plugging the fire hose of customers up to that is going to affect it from a performance perspective. Got it.

Dave Schappell: [00:48:43] And did, did Jeff disagree about this launch too?

Or did he, you know, w w how did, how did that work, or maybe not even in this example and other examples where he had differing opinions.

Rebecca Allen: [00:48:56] So Jeff was really awesome to work for, because, well, he has that laugh, which is awesome. Um, but also he's pretty technical. So, um, you didn't have to sit there and like, go, he's never going to understand what we're doing.

He always understood what we were doing. Um, but we had this agreements that a lot of the disagreements came down to. I was like, uh, this, this is gonna go bad on us, but I am not exactly certain how. And he was like, well, If you can't tell me exactly how let's do it and see trust ourselves to figure it out and, um, trust the customers to be really happy about what we're trying to do while we work that out.

Um, and sometimes I felt really strongly about the, this isn't going to go well. And, and, uh, and I'm a very stubborn person. And, you know, we had a couple of rounds of this. We had one on author interviews. We had one on this one and I, there were moments where I was crying in his office, but he was always really good about it.

He was like, he never meant to be quite that forceful. He always felt you always backed way off and was like, no, no, no, no, no, no. And it wasn't that I was saying, I wasn't going to do it. I always was willing to do. He was the boss. This is his company. I am a cog in a very large machine and I fully respect that.

But a lot of the pressure came from, he really wanted to convince people. And sometimes I was not going to be convinced. I be convinced when things turned out and the vast majority of the time, he was a hundred percent, right. In every respect, if there was a Bible, like in this case, it wasn't a matter of, he said it wasn't going to happen.

It was more a matter of are we going to be able to cope with it? And we totally did cope with it, although it did definitely cause some stress on unrelated people along the way. That's great.

Dave Schappell: [00:50:47] I had one, one, one other thing I wanted to ask you about it came up in the Dwayne Bowman and Ruben Ortega interview.

And I think you have some more to add to it. It had to do with why we had alphabetical search results, uh, in the very early days.

Rebecca Allen: [00:51:03] So Shelton a really, this isn't one of the, I think this will be in one of the blog posts you're posting with this before.

Dave Schappell: [00:51:10] But, uh, Rebecca did a favor, a big favor, and she wrote up a whole bunch of details about. The creation of the ASEN and, uh, and, and this topic as well. So, uh, those will be linked from the show notes and on the blog.

Rebecca Allen: [00:51:23] So show that and very clever solution to how are we gonna let people search by things like author and title and subject, and still have this thing run really, really fast and store in a small space.

So the indexes for author title and subject were also key value databases. These are very odd databases and that the value was empty. There was nothing there. They were just list of keys, right? And the key had some information in it. The key started out being a word from whatever you were searching. So at a word from the title, um, or an author, I think it was just the last names.

Um, and, uh, the third woman is a subject database. And so the, the first element in the key was that, so you'd get like, you know, database that like had, you know, everything from aardvarks does zebras with cats somewhere in the middle, at the beginning. But every cat entry, obviously there's going to be like 9 million titles with cat, right?

Just the, at cat who books. Right? So then the next element of the title was the first few letters of the first, you know, and letters of the title. So, you know, all those a cat there'll be a cat and then a cat who, you know, walks through walls or whatever. Uh, it won't be the whole title, but a chunk at the beginning.

And then finally there would be the ISB N is the third component. So that has to be ensured that each of those lines was unique. That's why the eyes being run was not stored in the value, because if you, well, you could have and just line them all up. And it, anyway, this was much clever it for technical reasons, it worked, it ran faster, right.

This way to look something up. So, um, as a result though, you only have one search order when you display the results on the page. And that was alphabetical by title.

Dave Schappell: [00:53:08] Got it. Well, that's, uh, it explains things at least. Um, and really at the time it was speed was probably the most important thing, given that people were on slow modems and, and it was miraculous anyway, you know, to have that massive database available for people to search.

And then they also, uh, Ruben and Dwayne also made the point that the search box wasn't like it is now, like, it was more common to have the multi field search. So you could type in an author's name, the title of the book, I don't know, maybe the category or something. So you could actually have some more precise searches that might make alphabetical results, not as bad as you would think as well.

Is that right?

Rebecca Allen: [00:53:49] Absolutely. So if you were doing a multi field search and again, I, these details are very hazy. I did nothing with the search engine, but my recollection is, is that the multi-feed search was, you know, you'd go do the title, word search, and then you peel off the yep. The one that matched all of them.

Dave Schappell: [00:54:06] So stepping back then from all the detail, I had two things I wanted to ask. One is the traditional, what's the advice for young entrepreneurs, you know, working on this type of stuff. But before that, did you ever think about how all of this catalog work and the catalog overall, how that helped set up Amazon competitively?

You know, because I've heard stories, we even read one yesterday that, uh, Alex Edelman sent to us he's he was an interview number six. Um, basically talking about how internally it really, I forget what the terminology was, but it set us up very well competitively with eBay, because we had structured data and unique products and title authority.

And like, were you thinking about that back then? Or was it more just thinking about it broadly, like a well engineered a solution?

Rebecca Allen: [00:54:53] It was a little of both. I mean, a lot of this is the result of a lot of people who think about things differently, but are very articulate and very intelligent working together.

This was very much an environment where people had. Well-developed opinions, state, you know, uh, there was really careful algorithm choice that sort of in the code, not in the catalog. Right. Um, but that support, so making sure that you can support a very large catalog is partly about how you store it and partly about what you do with it.

Right? So part of it's the data, part of it's the logic. We had people that were capable of thinking about both of those and thinking about how that was going to scale and also thinking about how can you sort of get as much out of the data that we had. We had, you know, we had data that was used for one purpose that we were using for another purpose designed for one purpose, using it for another.

And that was, um, somewhat structured. We tried to preserve structure whenever we had it, we tried to add value wherever we could. And a lot of this adds up to resilience, right? Because if you have different kinds of people worrying about different things and. And optimizing for a business where everybody's planning on scale.

Everybody's talking about, even at the very beginning, even long before Jeff was actually telling any of us, he was planning on doing all these products. We were all like thinking in terms of all these products, because we could see how this could be repurposed to sell.

Dave Schappell: [00:56:19] Yeah. I can imagine you all sitting there was like, well, what happens when we saw music?

I mean, that was just a natural discussion. Right. And then, and then some extremely big thinkers, like what happens when we start selling cement and how will that work?

Rebecca Allen: [00:56:31] Yeah. I don't know that I ever thought I'm cement, but I definitely spent a lot of time thinking about like clothes.

Dave Schappell: [00:56:35] Yeah. Well, that's awesome. Um, I always ask about at the end, like what's the advice and it's not necessarily just for entrepreneurs, but really, you know, people that are dealing with similar catalog issues and it, it probably exists with Airbnb. You know, if they're thinking about how do we catalog apartments and then houses and then cars and, you know, uh, you know, so what would the advice be for you for a company that's growing really fast and, you know, thinking about something like a catalog, you know, what do you, what would you tell them to think about, or, or try and mistakes not to make?

Rebecca Allen: [00:57:09] So that's a hard question because the technology has changed a lot. Yeah. Right. The cloud didn't really exist back then. So, um, there are a lot of options that are available now that were not available to us. Then I guess the thing that I think the thing that I would focus on is. Um, think about your algorithms.

If you're planning on growing, you need to make sure that you have algorithms that will stay fast, even with very, very large numbers of everything you can think about.

Dave Schappell: [00:57:33] One thing that I heard through a lot of your answers is like, there's bias for action, but then there's really think deeply about what you're doing, because it's going to have ripple.

And I don't know how you would phrase that, but it's, you know, it's like, there's so much careful planning that has to go into something like catalog because it's kind of permanent once, you know, you make that change or you add that field or, uh, you know, I don't know how you'd phrase that, but it's sort of like careful bias, right?

Rebecca Allen: [00:58:00] When go live, you don't want to destroy the story. Right. So I guess the way that I would think about it is I'm a huge believer in bias to action. I was never slow down. We're doing it. That was not me. I am a very, I'm a quick and dirty person, right? I am not going to refine that thing. I'm going to roll it out and then see what needs to be fixed.

So I'm very much in favor of that, but I'm also, I guess, the best way to say it is. From a wilderness first responder class that I took with another Amazon, um, person. And, uh, that is, don't just do something stand there. Right? Take that moment when you approach a scene and take it in, make sure you understand what you're looking at.

Know your problem. If you know your problem, your solution will be better.

Dave Schappell: [00:58:46] Yeah. Well, that's awesome. Um, thank you so much for being my guest on the podcast. Uh, for the listener, you might notice a little change, cause we actually had to record this over two days due to a, uh, an internet outage, but, um, uh, it, it, it worked, it seemed to have worked.

Um, but I think listeners are really going to like all the detail here. And thank you for writing the two, uh, the two blog posts that go into a lot more detail about some of the, you know, ASEN space, 36, like I actually learned, uh, quite a few things. So, um, it's, it's really impressive that, and like I said before, it's really impressive that the things you developed.

You know, it's 20 years now. And then, you know, and Asians are still holding strong at B zero eight or so, you know, so yeah, there's some runway there.

Rebecca Allen: [00:59:36] It's still day zero day one or day 1.0. Right. Who knows what's coming. So thank you very much for doing this. It was a lot of fun to think back over what we did and why we did it and get a chance to tell some of these stories.

This was an amazing group of people to work with. I, I wouldn't have changed a single thing.

Dave Schappell: [01:00:01] That's awesome. Well, for the audience, thank you for listening to the invent, like an owner podcast. If you'd like more details about what we discussed today, or you want to check out the posts that Rebecca wrote that have a lot more detail, um, please visit [invent like an owner.com](https://inventlikeanowner.com) and when you're there sign up for the weekly newsletter.

So you get notified about future episodes. Uh, that's it for today. And remember no sniveling.