

>> As Social Security takes on the challenge of baby boomer retirement, it's trying all sorts of strategies to become more efficient and more in tune with the first fully online generation. Among the many initiatives, three affect SSA's technology shop. Chief Information Officer Robert Klopp explained what they are.

>> Accelerating the use of data-driven decision making, transforming the IT investment process, and establishing program management office. And really those second two are more about how we govern the IT process. And this first one, which I think is the most significant one around, data-driven decision making is really where we're trying to improve the technology that we make available to the agency.

>> Let's start with data-driven decision making itself. What does that mean in the context of Social Security?

>> Sure. So, you know, the way to think about this is, it's sort of three things that kind of have to happen in this order. The first thing is, you know, we have, as you can imagine, lots of data in the agency. We've got, you know, what we call enumeration data, which is all the data that we – where we capture the fact that people are born and give them Social Security numbers. We have benefit data that talks about the information that we need in order to determine whether you're eligible for certain benefits. And in the case of SSI, we have income data that helps us understand how much money you have put away for your retirement and how we're going to pay that out. And then, as you can imagine, there's, you know, within those topics, there's just terabytes and terabytes or even petabytes of data associated with all that. All of that data is personal information. And so we're very careful about how we store it and how we use it. We

have very strict policy guidelines about, for almost all of this data, about how things can be used. So this first part of the problem is, how do I take all of this information out of the kind of operational systems that we run day in and day out in order to make stuff go and put that information into sort of an integrated format where it's extremely easy to query it and ask questions and get answers back? And the technologies or – that we would use in order to do that are basically lumped into a category called data warehousing. And so we're building an enterprise-wide data warehouse that's going to take a very broad swath of that data and integrate it together so that it's all in one place, it's very easy to ask questions of it. And, importantly, it's all in one place with controls over it so that we can ensure that, first off, no one can get at it from the outside; bad guys. But second off, so that when we query it in order to provide answers to people, that we're not violating any policy around the use of personal information. The second part of the problem, once I've got the data assembled, is to provide tools to allow people to ask these questions in a meaningful way. And we're really looking at three sets of tools. Again, the vernacular of data warehousing one set of tools is called business intelligence tools, which is, I mean, there's lots of bad jokes about business intelligence being an oxymoron, but I won't go there. But business intelligence tools are tools that allow you to query the data using sort of either natural language-ish things or other kinds of online query tool capabilities and get out basic reports. You can ask very sophisticated questions and get very sophisticated reports, but the report formatting would look like reports, or they would be simple graphical things like maybe pie charts and bar charts, the kinds of things that you would, you know, might typically do using something like Excel, only instead of tapped into a small

amount of data in a spreadsheet, these tools will allow you to tap into terabytes of data in a data warehouse.

>> We're speaking with Rob Klopp, he's the CIO of Social Security. And a couple of questions just backing up to the warehouse for just a moment. Would it be that the warehouse might be a smaller dataset than the – that is, the dataset from which you withdraw the warehouse information, because you would remove, say, personally identifiable information connected with that data so that you could look at groups of people and so forth without finding individuals?

>> So it's an interesting question. First off, the data warehouse is almost always significantly larger, and the reason for that is because we not only store the information but we store a history of how that information has changed over time. And so that allows us not just to ask questions about groups of people, but actually ask questions about trends across those groups of people. The second thing is that typically what you do in a data warehouse is you actually store the raw personal information, but then you put, let's say, policies in place that allow people to ask questions that obscures the personal information. So, for example, there may be a record in there that has Rob Klopp in it, but if I ask a question and say tell me about, you know, the average, you know, age of people that have worked for federal government, or something like that, the answer now comes back with the personal information aggregated out. So the data warehouse itself probably stores all of the PII at a detailed level, but we put filters on top so that we aggregate that personal information out whenever you ask a question.

>> And do you have any examples or do you have any queries already from executives at Social Security as to the type of questions they would like to ask and get answers to that maybe it's really hard to now in the absence of having this data warehouse and these tools?

>> Sure. I'll give you a great example of something that we're starting to work on now that's got everybody very excited, because even though when I say this, it will sound like a very sensible and straightforward business question, we've never been able to integrate the data in order to be able to get an answer in a meaningful way. And that is, we have currently one sort of silo of data that tells us about all of the people that come into our field offices. We have another silo of data that talks about all of the notices that we send in the mail to people. And we know anecdotally that some of the language in some of the notices is very obscure and it causes confusion. But we don't have the ability to see explicitly the connection between sending people a notice and driving them into the field office in order to get someone else to explain to them what the notice says. And so for the first time ever we're going to be able to connect the fact that when a particular type of notice goes out, that the people that receive that particular type of notice have an increased likelihood of now trying to schedule a visit into a field office to have someone explain to them what it really means. And when we start to see those connections, it will allow us to go focus on these notices that are confusing and find ways to rewrite them to take the confusion out.

>> So this is something the IRS has been working with for a number of years that they find that the form is the problem and not that people are stupid and can't figure it out.

>> Right. And when we think that that's what the IRS has been doing, which is about simplifying the language generally, that's something that we're working on. But my point is that more than just a general attempt to try to make the language more consumable by the public, that we're actually going to be able to identify what language is causing the most angst and let us focus our attention on those things.

>> And what about the actuaries that look at the long-term affairs of Social Security? Is there anything that they could be boosted in their decision making with this kind of warehouse you're building?

>> Absolutely. They're one of the people that are probably most excited about the opportunity to go use this data. You know, the fact is these guys are data geeks by definition if they're actuaries, so they're very used to data-driven decision making. And, you know, I'll be perfectly honest and say, you know, I haven't had specific discussions with them about how they're going to use the data, but I will tell you they're probably our biggest fan in the agency around this particular critical eight initiative.

>> All right. So tell us, then, how to get to the point of where the warehouse is operational and working. What do you need to do between now and then?

>> So we've been working on this for a while. By the way, one of the things that's interesting is we're going to build this data warehouse in the cloud. So this is –

>> You're hitting all the right buttons.

>> That's right. We're hitting all the right buttons. This is the most cost effective way to do this. There's lots of distinct advantages for how to do it. It allows us to scale the compute as big as we need to scale it in order to allow the answers to these questions to come back quickly. So we're moving very aggressively on this. In federal speak, we need to get an authorization to operate before we can take and put personal information on any new platform. And we are, you know, we just got our first feedback from our security folks on the kinds of things we need to do to go tighten this up in order to get this authorization to operate. And there was no real show stoppers in there, so we're starting to work through that. And our plan right now is to have a data warehouse populated with tens of terabytes of data, actually I think we're shooting for about 40 terabytes of data by the end of this calendar year, and we'll start rolling it out to sort of beta test users in the July timeframe.

>> And by cloud, do you mean commercial cloud?

>> I mean commercial cloud. We're going to actually do this one in Amazon Web Services.

>> And so is this a FedRAMP situation or do you need more security even than that would permit?

>> Well, it's FedRAMP. But FedRAMP, you know, is only part of the security problem. Being FedRAMP certified doesn't mean that it's perfectly secure. So on top of FedRAMP, we ask our cyber guys to take a look at what's going on to provide us this ATO, or authorization to operate, and it's that very specific detailed cyber analysis that we're going through right now.

>> Social Security Administration CIO Robert Klopp. Tomorrow in part two of my interview, Klopp describes how he's revamped SSA's IT investment strategy. Share this interview with colleagues. Go to FederalNewsRadio.com/FederalDrive and click "Share with friends."