2015 Air and Space Conference

Empowering Tomorrow's Analyst: Revolutions In Analytics

14 September 2015

SPEAKER: The title of this forum is Empowering Tomorrow's Analyst, Revolutions in Analytics. Today's Panel will focus on how the military and industry, are embracing the onset of the information revolution and its impact on the Air Force ISR Enterprise.

Our Panelists include Lieutenant General Robert Otto, Deputy Chief of Staff of Intelligence Surveillance and Reconnaissance; Dr. Steven Rogers, Air Force Senior Scientist for Automatic Target Recognition and Sensor Fusion; Samuel Druker, Director of Data Science, Windows & Devices Group, Microsoft; and James Crawford, CEO and Founder, Orbital Insight.

Each will make a short presentation and we'll open it up for questions. We will have a wandering mic, so just raise your hand or stand up and we'll get the mic to you, and then you'll be able to ask your question directly. Gentlemen?

LT GENERAL OTTO: How about now. That's working. Okay. Hey, thanks. I appreciate you all being here. We've got until about 1555 to talk, so we are not going to talk long, we just want to warm the appetite for the subject matter. And I'm glad to see so many military people in here because what I wanted to do with the panelists was to open our eyes, is the military to some opportunities, so the backdrop on this is, a couple of years ago, we produced a survey from the Headquarters' Air Force, and we sent it out mainly to the leadership and we asked, how would we do if we had to do analysis in an anti-access or a denial situation? And the respondents seem to indicate overwhelmingly is, not so well.

And, you know, for reasons that might imagine about 75 percent of the Air Force now has come in after 9/11. They are very good at doing what is being asked of them, for counterterrorism and for counterinsurgency operations. But they haven't been focused on the sorts of challenges that we would face in an A2/AD environment.

And then you add to that, our distributed Ground System which is the best on the planet. If you think of a regionally-focused, globally networked, immediate war fighter support, it's exquisite, but it's millions of lines of code, and it takes a great deal of work to deploy or integrate new sensors throughout that weapon system. And then you look at big data, everyone wants to talk about big data, you now see it on advertisements everywhere, what are we going to do with big data?

And we haven't really come to grips with that within the military. And then you add on to that, just data transfer rate, so when I was out at Beale, it wasn't uncommon to see data transfer rates in the five megabits per second range, and now you are looking at things that exceed or come up to 3,000 megabits per second, so there's a lot of data that come through the pipes and add to that, some of the new sensors that we have, like Gorgon Stare, which is wide-area imagery, motion imagery, or LIDAR, and a whole host of new signals intelligence sensors.

And then you add on to that, and I just did a California swing; there's so much happening in the commercial space. And, which, you know, we are not talking about renting warehouse space; we are talking about in space, commercial applications and commercial satellites. And then finally, there is open source, and a plethora of reporting through open source. Just this last weekend if you -- humorously there was a Russian wife who was commenting about Russian Troops in Syria, and how she wasn't going to see her husband for eight months.

There's pretty good intel you can get from open source, if you know how to harness it. So, the question is, how do we develop analytic capabilities to remain ahead of a near-peer adversary, and I would specify with less resources, that we have today? And there is just four things that I'll comment on briefly, and then hand the microphone over.

The first one is, we've got to wrestle with access, so we've talked about the intelligence communities, information technology environment, which

in theory is going to open up data, and make it accessible throughout the intelligence community, but that's only true if the data is being shared, and there are some barriers to entry there that we need to overcome. But even if we get that, the second thing is going to be data flows to which I would say, we are working within the distributed Ground System on an open architecture.

Why? So that we can integrate new sensors more quickly and we are also working on multi domain, so new sensors from space, new sensors from the air, new sensors in cyberspace so that the data is available to the analyst. We need to look at open source as a source of data that gets integrated into our distributed ground systems. We need to partner with industry, and that's why Sam Druker and Jim Crawford are here.

I've spoken with both of their companies, Microsoft and Orbital Insight, and it's just tremendous the work that they can bring to the table. And the question for us is when do we lead as the

government, and the Air Force, and when should we follow? And I think that that is a question upon which a lot of money is wrapped up in, but also opportunity and we've got to get that right.

And then the last thing under data flows is, commercial and government tools. So we have tools in the government that may be NGA developed, how do we make those tools accessible within the Air Force but the commercial sector is going to have tools, some of which are phenomenal, and in my travels I've seen some that we really need, but how do you know which ones, how do you vet them, and then how do you bring them in on the timing that you need.

And that gets to the next point, which is enabling technology, and that’s the reason that Dr. Rogers is here to talk to that. But, you know, if you think about technology as the ways to ease the manpower burden, then there's some real opportunities there. Just one example, I was talking to a young -- a senior airman out in California on Thursday and she was a screener, typing madly in the middle of an

operation, and I just said, how would you like to have speech to text, where you could just talk it out?

And she laughingly said, oh, that will be great. And I said, well, yeah, it's coming soon to a theatre near you, we are about 97 percent there. And then she was blown away, you are serious. And you could see her eyes just light up. There's stuff that we can to provide tools to our analysts, that will make a big difference, and again, with the goal of easing the manpower burdens, we can do some of this other analysis.

And one of the things we have to get right as a government we ran a Plug-Fest Plus where we ran a little competition, anyone could enter it, and then we were going to lay out a contract within 60 days. Well, we learned some things, we didn't get the contract out in 60 days, it probably ended up being more like 75 or 80 days, but the goal is still in weeks, not months and years. And we have to do that, I think, if we are going to be able to harness innovation.

And then the final thing is training, and of course we are going to have a ton of work to do from a training standpoint, because analytics in the future are going to be very different than analytics today, and so we have to focus on that. That's why we formed a wing, the 363rd Analysis and Targeting, is their focus. The ability to rapidly retrieve and assess and act, kind of indiscriminate of the sensor or the domain, is the goal, and if we can do that, we can inside of the enemy's OODA loop and kick their ass. Okay. Over to Dr. Rogers.

DR. ROGERS: So I'm a geek. You know, I'm a machine-learning geek by trade, and I think that's why I'm here, but I'm never really sure. Sometimes I guess I'm just the comedic interaction here. But I've got to tell you, the challenges are immense, but I've never been more excited, this is a really exciting time to be a geek, we are making some amazing breakthroughs. And I'm very optimistic, I'm the most optimistic I've ever been. And my optimism doesn't lie in the fact that we've already solved the problem,

or even that we know how to solve it.

It isn't that computers are becoming faster. I don’t believe any existing algorithms or set of algorithms, or any existing approach to computing are going to solve these problems but our advancements are becoming so quick I'm really very optimistic. But I do believe -- Jeff Jonas, for example, from IBM has a graph he shows which shows the amount of data available is growing exponentially, but the [sensor-making] ability is leveling off.

That’s the problem for people like me, how to close that gap. And again, we are making inroads, but we don’t know how to solve that yet, and we are not going to do it by ourselves, it's going to be a bunch of commercial breakthroughs that are going to help us to do that. If you look at Syria, you look at Watson, and you look at the lessons we've learned as you look at those approaches, and you see where they failed, and then you try to map them into our space and you realize, hey, there's some really important military specific issues that we face that, maybe, the

commercial world doesn’t face.

Speech to text that General Otto brought up is a great example, everybody has speech to texts on their phone, but if you try to put into one of our environments, it just doesn’t work. It doesn’t help our analysts, it actually slows them down. So we have to have a lot of secret sauce, if you will, to make those sorts of commercial breakthroughs viable for our mission needs. We know that, we are figuring out that through that we learn lots of lessons every day.

Big data, another spectacular example, the idea of 33 analysts, again, doesn’t work in our domain. We can't afford to not understand the correlations that the big data analysis is finding. If we just use them blindly we won't know when not to use them, and we can't afford those mistakes. So you keep following those down, and what you do is you look to the lessons that the commercial world is teaching us, and what you find out it's not an algorithm, it's not a computer it's an ecosystem, and that’s why we are all in on just fundamentally changing the

ecosystem of our intel.

And if you look at the open architecture work that General Otto just pointed out, that is a spectacular example. It won't be a one-shot thing. Look at Apple, they just introduced the next iPhone, that's the hardware and software layer that our open architecture work is going to provide, or DCGS, but that will be -- That happened to us more than once, and then we are going to build the layers on top of that, to allow the innovation of commercial. So that's going to happen.

So now I have to ask you the question, because we are going to use more and more commercial breakthroughs, both on hardware and software, and in fact we are going to rely more and more on commercial data, open source data, space data. So the question becomes, General Otto's challenges, how do I provide you, the war fighters, decision advantage over our adversaries, when we are relying on commercial products and commercial data. That's an interesting question.

I think the answer lies in trade craft. You know, what we do in the intel business is really unique. How we interface humans and computers together to do what we do, or magically do is unique, and that's where our advantage will lie. It won't be in the technology, it won't be in the sources of data, it will be how we do that trade craft.

MR. DRUKER: My name is Sam Druker, I'm here from Microsoft Corporation. And within Microsoft, I run a Data Science and Machine Learning Analyst Team, it's central to the production of Windows, so it's a big summer for us just launching Windows 10. We were traded at the inception of Windows 10 about two years ago, at the start there were three of us who started the team, or a team of about 150 analysts that fully support the development and marketing and production of Windows for the phone, PC, laptop, tablet, X-Box and services.

My experience is as at Commercial Practitioner of Analytics. Windows is a pretty widely-used operating system. We have about a billion

devices in the world, that [inaudible] every month over thousands of properties in the consumer market. And we have taken a brand new approach, in the last two years, to how we use that data to inform the product development the servicing, and operation for windows.

It's been a crazy kind of journey, Windows has been around for a long time and it has a lot of really entrenched way of using telemetry and data. And really the story of last few years for us has been about how we changed decision-making, changed our culture to use modern big data techniques. How we dealt with things like getting all of the various parts of windows, and indeed all the divisions of Microsoft to share data, which seems very analogous to sharing data in a lot of other environments, and getting it pulled in, into a common place that we could actually do analysis on a regular basis.

In the two years' time the team has produced about 467 insight projects, and analytic projects that we have measured impact for in the product in

operation of Windows. We've grown the team quite a bit, and we've come to understand a lot more about what it takes to run a pretty large-scale commercial data science machine-learning operation.

That’s included advances in technology, but also things like infrastructure, sourcing talent acquisition, as well as process, and really some of the core values. We ran into some of the things, that I think everybody runs into here, things about common access to data, really isn't much prevue for analysis if very body can't get to the data. Another thing that’s super-instrumental for us is understanding that peer review was very important for correct analysis. The methods that we use are actually pretty tricky, to understand the pretty leading edge, and they require a combination of talents that very few single individuals posses.

You need to have a background in computer science and statistics, and modeling as well as domain expertise to add to the mix. And so having a good peer review process, having a good team process was

critical to our success. We worked a lot on things like data quality and infrastructure to present data that was easily consumable, whether that data was generated from our own instrumentation, or using other public sources.

Things like reverse eMaps, for networks, for country codes, or foreign codes, as well as other sources of data. We also worked very hard to make sure that we had schema in modeling for data across the various domains and disciplines. And this is one of the more fruitful sources of productivity gains; understanding that when we do something like counter interactions on advice that we counted consistently in a way no matter whether the applications are being used, for example, when we had support calls, having a way to map that into things that actually happen with the devices.

Dr. Rogers mentioned Watson which is, of course, one of the oldest telemetry systems in windows. Revamping that system and getting to as close to an air free state as we probably ever had in

the Windows kernel has been also begin to shift. So we've been through that process. Of course there are other parts of Microsoft that build some great tools for machine leaning and analytics, starting Excel and going after things like Azure ML in the new Proton analytics platform.

We are early adaptors of almost all of that platform, as well as many open source and other third party solutions. So, our job is to actually do the analysis and get it applied commercially, and that's what we focus on with that.

MR CRAWFORD: Good afternoon. I'm Jimmy Crawford, from Orbital Insight. We are a small startup in Silicon Valley, we are interested however in very big problems, we are interested in how to do very large-scale analysis of satellite images, and by large scale we usually mean, millions of images at a time.

So picking up from something General Otto said, from a commercial space point of view, we live in very interesting times, for years, decades in fact,

launching a satellite has been 3, $400 million endeavor. Nowadays the guys in plant labs, they'll launch them 100 at a time, and they are little more than cell phones with a big lens on the front, and some solar panels on the side. But they are still able to get 5-meter imagery, and we see a day in the not too distant future, between Planet Labs, and Urthecast, and Sky Blocks and Black and other startups where we will be able to get daily coverage of every spot on earth through commercial space.

The interesting thing from my point of view, I'm a software guy, I've always been a software guy. I used to run the Google Books Project were we scanned 20 million books. We took picture of every page of 20 million books. We put it all through OCR, we figured out what the words were. We put it up on Google Search. If you type in "to be or not be", you get a picture of Hamlet.

So the fundamental insight behind Orbital Insight, is why not do the same kind of pipeline we did for Google Books, except for satellite imagery.

Because in the future you are not going to be able to have people interpreting satellite images when you have them coming in basically at every square kilometer on the planet every day. So, that's what we are doing, we've gotten fantastic support from the VC community in Silicon Valley, and built up a team.

So we went to Wall Street about a-year-and-a-half ago, and we were able to talk to 20 of the smartest investors on Wall Street. And we asked them, so if you could measure anything about the earth that you currently know, that's vaguely visible, what do you want to know? And they said, well, we think the price of oil is not completely stable. And keep in mind this was a-year-and-a-half ago, so think it might go down but we are not sure.

And the problem is we don't know how much oil there is in the world. And the interesting thing from a satellite point of view, is that oil, crude oil is stored in tanks with floating lids. So the lid of the tank sits right on top of the oil. And they have to do that, because crude oil contains propane and

various gases components as well as tar and oil. And they don’t want the gaseous parts to get out, so they float the lid right on the oil.

So from the satellite point of view we see these lids go up and down, we can see the shadows of the edges of the tank cast on the lid. And keep in mind these tanks are huge. Two of these will hold all the oil in a supertanker. But there's -- about 17,000 of them around the world, they are owned by different companies, they are in different countries. And together, those 17,000 tanks represent the sum total of supply and demand for oil. So when the U.S. starts fracking more oil goes in, when China's economy less oil goes out. Eventually a surplus builds up, and suddenly we have $2 gas at the gas pump.

But if you can see the tanks, it's not sudden. The reason it seems sudden is nobody knows the total of all those tanks around the world. So we've been putting in place, following the advice of our friends on Wall Street, the ability of satellites to track 17,000 oil tanks around the world, and see how

much oil is in the sum total of them.

Similarly we can track all the cars in all the Walmart parking lots and say whether Walmart is going to have a good quarter. The other thing the Wall Street guys told us, is they thought the Chinese economy might not be completely stable, and keep in mind and so we started tracking construction in China to see whether or not it was slowing down.

So, overall it's been a fantastic experience, we are about two years old now, and growing very fast, and we are very excited and optimistic about what can be done in the future, not just to understand things of commercial importance but also for government applications for NGOs, we've got a really great project with the WRI to try to predict deforestation before it occurs, a project with the world bank to try to map poverty around the world. So, I think this kind of technology any number of applications as we start to figure out what to do with this increasing flood of satellite imagery.

SPEAKER: So, questions? Over here, can we

get the mic there? Coming your way.

QUESTIONER: Thank you. A question for General Otto, and Dr. Rogers; you talked about speech to text and how the revolution is coming, 97 percent there, and some of the challenges of making it work for the analyst. So back to 2005, Air War College, I wrote 100 percent of my papers using speech to text, using commercial Off-The-Shelf, Nuance, Dragon, and that was an offshoot of an earlier DARPA project, [inaudible] assistance, super committing that had failed, but that was a great technology that came out.

The question was -- Well, a little more precursor, there was some training, and then there was some teaching, what's a JFAC, what do I mean by CFRS? It had to work, but after you got through that learning and training curve, it worked fabulously. You drive down the road, 20 minutes later page, click File, click Save, and the thumb drive to the teacher.

So the question is, if it worked that well then, 10 years or 9 years now is a long time for it to get to the analyst. What were the barriers? Is it

just technology, because it seems that technology -- or is it more operational? Is it integrating? Can you talk about what took so long?

DR. ROGERS: I won't presume to explain the processes that were inhibiting the transitions and technology. On the technology side you hit it. I mean, you can take Nuance or Dragon, or any of the other solutions, they will give you some level of accuracy, the problem is if you take them off the shelf, the vocabularies are not the right vocabularies as you pointed out, and you will have some [inaudible] out there. And so by the time the analyst had to make the corrections, they could have easily have typed it in.

So just from a manpower perspective, you know, if you ignore processes, from a manpower perspective it wasn't advantageous for us to transfer that out of the box. So it needed some change, it needed some middleware, it needed some wrappers around it, we are doing that, and we are doing the transition now. Could it have been done sooner? If there had

been a driver, if the processes had been right, we could argue that point. Okay. So more demand on the analyst time meant let's reach for that and let's pull that in. Okay. Thank you.

LT GENERAL OTTO: Yeah. And I'll just add to that, that's exactly the kind of an example that we want to do better on in the future, because we probably could have done a Plug-Fest Plus. Okay, we know that we have a commercially viable product, and we also know that we are going to have to make some changes to it, in order to make it viable for us. Let's send out a message to the entire commercial activity and say, hey, who thinks they can integrate what a MAM is, or any of the other acronyms, and also the ability to perfect it from the screener's point of view when they see an error.

You see a consistent error and then somebody can update it across the enterprise and we could have been here a lot sooner, would be my hope, and that's certainly what we envision for the future, is to be able to turn this kind of stuff around. And part of

the question is, you know, where are we going to see the bang for the buck.

And one of the things that both Sam brought up, and Jim Crawford, you know, you look at the teams that are figuring out, how do we go after this problem? And this is not talent that is commonly held in a single organization in the Air Force. I don't know. Sam, would you be willing to talk about, so your Data Science Team, how big is the team and what kind of specialties they have and so on?

MR. DRUKER: Sure. So we started the team about two years ago, there were four of us that founded the team; it was new to Windows at the time. We support an overall engineering effort of about -- I think the number is about 12,000 engineers involved in the development of the operating system, and really the parts. And we found that what we needed was, for each project we needed a range of talents across traditional software engineering, formal mathematics, statistics probability, some experience with the practical tools for big data analysis; whether it's

stuff like Hive and Hadoop, or other real-time solutions.

But we also needed some domain expertise for each of the problem spaces that we attack, working on an upgrade problem, working on a print problem, different -- working on a kernel code. So, what we did is we decided that we would treat it like a multidisciplinary discipline. We brought in folks who have different kinds of formal training in some form of numeric methods, it might be simulation from astrophysics, it might be biomedical background, it might be data modeling, we have folks who have come from an actuarial background, user and market research backgrounds.

And we line them up, side-by-side, where people have these other skills. We formed teams of two to three for each of the projects that we undertake. We've had a great deal of success bringing in folks from the college recruiting program. Microsoft tradition has a pretty strong college recruiting program, but we designed a section of it

specifically to go after data science and machine learning talent. And we found many places that had many of these bullets, kind of covered off, but needed a way to bring them into working on real-world problems.

One of the things we also recognize is that this is an apprenticeship set of skills, you don't walk out of a set of books, and you start applying it. You know, it's requisite for the fundamentals but we needed to bring people along and work on various projects with different kinds of scope. And then the last thing I'd say is that a large part of the predictor of success for any given project, for us it's been figuring out the right framing of the problem to go after.

Because of course the framing provides the context that makes a ton of difference. You know, speech recognition for dictation of a thesis paper is a lot different than, say, speech dictation you might use for a civilian example the 911 Dispatch, for example. So recognizing the framing needs to change,

and that's where only people with domain expertise can really bring that in, and you made reference to a trade craft before, and that's something we have seen in the equivalent the equivalent in our world.

SPEAKER: The next question? Over here, Ross?

QUESTIONER: Thank you. Very interesting presentation. In one of the previous panels, one of the speakers talked about how the Chinese have been doing a tremendous amount of writing, about information superiority, and one of the things that, DoD has run into is they don't have the resources to translate everything that the Chinese are writing in any given year. There is that much information just in that one tiny domain.

So, here you are talking about photographing every square meter, or every square kilometer of, I presume, the land surface, or maybe not just the land surface of the earth, every day. And we have heard other people talk about this flood of information that that is coming. So what's the strategy, what's the

approach not to throttle the pipe, because that's clearly not the right answer, but to find a way to pull the real meaningful information out of that vast flow of data that's coming the analyst's way?

MR. DRUKER: So, I don't know if we have the final answer, but I can tell you how we think about it. So, we think about this as a two-step process, the first one I call pixels to numbers. So it's taking a picture and counting whatever it is that you care about. Whether it's the number of pixels that are in shadow, which gives you an idea of how tall the buildings are. The number of chlorophyll atoms in the cornfield which helps you to predict cornmeal. The number of airplanes in a military base, the number of cars parked at a research facility. All these things are pixels to numbers.

And then the second step is numbers to insight. And the numbers to insight goes right back to the question that Sam was just answering, and it's all about data science, because the numbers are just a signal, and that signal contains a huge amount of

noise, or I should say, that signal is hidden within a huge amount of noise, because we are talking about the real world.

So everything we do involves equal parts machine vision and data science, and that's what takes you from an underlying picture, up to some insight. Now having said all of that, I think the humans are always really very important and in some sense what you are looking for is generally is a needle in a haystack. And what we are trying to do with the automation is to sift out the haystack into a very, very small pile of hay so that the humans don't have very much to look through.

So, the third part of this is the tipping and queuing for the humans, so if the humans can spend their much more valuable time looking at a smaller number of images, which are the ones that really matter.

So, lastly before I pass on the mic, you are talking about the Chinese and the data competition. There's an interesting recent story that happened in

my world of deep learning and machine vision, where on the internationally-recognized best benchmark for machine vision, a few months ago there was a serious competition between Google and Baidu, and they kept passing each other in terms of accuracy, until they were both getting down to the point where they were basically as good as humans, on this particular corpus, which was a huge corpus of images, and hard decision problems. Like telling one breed of dog from another, and not telling a dog from a tree.

And in the end the Baidu guys were disqualified because it was judged by the international group that was running the competition that they had cheated by submitting too many submissions in too short a period of time into the contest computer. But other than that, Google and Baidu were running neck and neck on what is, in some sense, the first key problem and image interpretation. So I think the question you ask is very real.

SPEAKER: Next, over here?

QUESTIONER: We are describing a situation

where we were able to do some exquisite analysis of big data, various sources and various types in all different kinds of situations, which implies to me a very adaptable analytical capability, where, every time we go into a new problem set we might need to develop some new tools, or new databases, or new methods of extraction. How do we need to restructure our talent pool within the ISR community to be able to do that kind of adaptation?

LT GENERAL OTTO: So, as you heard Sam talk about what goes into data science, the first I think is, we are not going to have that kind of talent. It's probably not going to exist within the civilian workforce; it's certainly not going to exist within our enlisted workforce that makes up the backbone of what we do. And then do we need to have that talent? Or can we access the commercial sector, to think through some of these problems?

I'll just throw this one out. You know, we've been battling since 9/11 and we are still using the number of hours that fly an MQ-9 or an RQ-4 as our

measure of merit, and we have not gotten our arms around, are we effect with this stuff? Because we've sure had a lot of it, we've gone from 1 CAP to 5 CAPs to 11 CAPs to 35 CAPs, to now we are at -- we were at 65, we are down to 60. Is it doing any good?

And no one can tell us that answer. And that's an issue and an indictment at several levels. At the Combatant Command level, at the OAC level and at the half level, you know, me. But we haven't been able to figure that out. I bet you if this was Sam Druker's problem, he would have had it figured out by now.

And so I think that there are some things that we can look to the commercial sector to wrestle with for us, but beyond that, it does imply a change in our analytic workforce. One of the struggles we have today is we are taking a very bright Airmen, and in some cases, and I'll talk about the four motion video example, we are having them do some pretty mundane stuff, and then those geospatial analysts get bored with it, and the additional benefit we provide

them is, you are doing it on shift work, and we are going to work for really long hours. So there's not -- That's a difficult problem.

If we told them to go solve problems where they could access different databases and come up with their analytic insights applied to a key intelligence question, we are much more likely to harness their interests and therefore retain them over time. So we need to change what the analytic workforce looks like. I see in the future, probably more all source analysts rather than the huge number of geospatial analysts that we have now.

Again, if we are successful from what Jim is talking about, going from pixels to numbers and numbers to insight, then it can change the game in the way that we do this. So I see a huge opportunity, and the question is, you know, you can't, because we are a bureaucracy it's going to take us a couple of years to transition that workforce, and to get them retrained, and the trick will be how agile can we be in turning that workforce into something slightly different.

MR. DRUKER: Not to get on as soapbox too much, but one thing that's really crucial for how we've made our investments here, it is, analogously, we have a bunch of data science guys, we a bunch of really good resident geeks, but that's not really the most critical parts of the success of the program. The most critical thing was getting the other 12,000 engineers that work in this product to understand how they need, both to consume the work that was done within the central team, but more importantly how they can adapt their current job descriptions to have a better understanding of analysis in the first place.

And this will master a couple of things, one is that they need to have a -- I think of it as a literacy problem. We wanted our engineers to become literate in statistics and analysis techniques so that they could solve the simple problems on their own, and apply their own curiosity and thirst for innovation in a way that we were never going to be able to from our, you know, somewhat little team in the central bureaucracy.

The other thing that was super-important is that provided that common access and infrastructure, so that everybody could do analytics work, and of course in the normal course of doing their job, where they could take the curiosity for doing what they were doing and doing it better, and have access to those tools to pursue it.

And then the third thing that was really critical for us, and I really think we would have been sunk without this, it's we found a way to make that curiosity about how to do better, about how to measure success at every level of the organization, and that was supported by very strong directive from the top, our most senior leaders in the organization made it clear that, this is where we are going, this what we needed from everybody. And we were able to accomplish a lot more with those three elements in place, which I think might be somewhat analogous.

SPEAKER: The gentleman on the outside, there?

QUESTIONER: Yes I went and purchased a new

computer recently. I'm enjoying my Windows 10 experience. Thank you very much. But the decision I had as I purchased this new computer was the storage device, so I had to think through speed as well as size, so I made a decision, I went for Hybrid hard drive. And so, why am I telling you this story? Well, I have a 16-month-old daughter and so I carry a camera with me, everywhere, I take a photograph of every single thing she does, and every photo we store on that hard drive.

Now, at the moment it's okay, I've got plenty of storage, I'm not running, but in 12 months' time when I go to my wife and say, now we need to spend more money on this new computer, she's going to say, no, that’s not an option.

So, my question is, I think that is not a story that is unique to me, I think it's unique to a problem shared across the world in all sorts of enterprises. So what is happening in a very boring and very uninteresting aspects of data storage retrieval, meta data, and understanding what it might

take today, which might be of utility, of tomorrow, next week, 10 years, 20 years, 100 years?

MR. DRUKER: Sure I can take a stab. I think that Dr. Rogers mentioned the storage explosion. This is an idea that goes back to Jim Gray and some of the first transaction systems about, you know, the exponential growth of the availability of stories. The good news is that that's moving even faster than Moore's Law for processors. We have more and more local storage available and, you know, we've gone beyond the spinning disk, there are technology beyond Solid State disks, I'm not an expert in them, but I'm sure our friends at Intel, or some of the other Solid State manufacturers could talk about that.

But local storage is, of course, part of the problem. The key thing, too, is that very much like the other Insight's problems, you will probably take 500 pictures of, you know, your 14-month-old over the next, you know, couple of months maybe, more weeks. I tend to be a kind of a cameraholic myself. The question is, which of the 40 that are really the most

interesting, which of the 40 that you really wanted to go back to five years from now, 10 years from now, that you'll just kick yourself if you don’t have.

And that’s a problem that is susceptible to both analysis, and some of the more advanced techniques with things like machine learning and data retrieval. So, you have both the twin problems of, hey, just raw bits, where do you put them, but also there is very much an analytic problem here of, which of those are going to have the most meaning, and having that prediction. There are other analogous problems. Sometimes it takes 20, 30 shots, right? Well, they are basically the same shot, but 17 of them didn’t turn out, so how does the system automatically understand that those are all the same. Jim is probably a lot more expert at that part than I am.

LT GENERAL OTTO: Let me say something real quick.

SPEAKER: Sure. Go ahead.

LT GENERAL OTTO: In fact, your point is exactly the one we are facing. The automatic

generation of meta-data is the key that we have to crack that code. And when we crack that code, and Jim mentioned deep learning, for example, if I can automatically generate a natural language caption of an image or a video snippet, I can begin to crack meaning, which right now is a human intensive process. It's the only shot we have at cracking the human capital problem we face in ISR.

MR. CRAWFORD: Yeah. Just to highlight that briefly, I think that's exactly along the lines. The storage problem will likely solve itself. In Moore's Law, as Sam said, is moving even faster for hardware, will largely save us there, but the problem is, if you've to that many images how Do you know what I mean? Every find anything?

It's the same problem Google has with YouTube, there is, I don't know, I forget the number, like five years of imagery uploaded to YouTube every minute, or some ridiculous number like that, right. So, it's a tremendous number. How does anybody ever look at all that? And I think the technical solution

there, which is coming and actually is moving very fast, is this whole ideas of deep learning for image classification, and I think the ability as General Otto said bout transcription, the ability to actually automatic caption generation is in our fairly-near future.

SPEAKER: And we had one here in the front -- Well, you get the mike, and we'll end with the gentleman here in the front.

QUESTIONER: I think, my question is for Lt. General Otto. So, you know, we've talked about a lot of the analytics that is out there, and how potentially game-changing that could be especially with, as we pull up some of the mega data, and thing like that, and will enable us to employ our air power better in the future. Now, when I look at it from the enemy's perspective, my question is, what does that mean for us, as we go to employ data, on the risks and vulnerabilities that have?

Do we need a counter-analytic strategy? Do we need some sort of counter-ISR thing, because if we

can obviously image every five meters of the earth 24 hours a day, or less than that on a repeat rate how -- that's going to present problems to us in the future, I think, so how do we get after that?

LT GENERAL OTTO: I'm a big fan of diversity, and you know, depth, and so one of the things about anyone that wants to tackle our Air Force, is we've got a number of weapons with which we can really put a hurt on somebody. And so it becomes holistically a problem that no one wants to take one, and that forms deterrence, and that's great for our country. I think the same thing will apply in terms of analytics, you know, the debate is do you put all of your eggs in one basket, which is then attractive for attack, or do you diversify it?

What the commercial sector has done is, you know, it seems to me, as a customer, because I just get on the cloud and my stuff is there, but unbeknownst to me it's backed up and imaged, and there is no way that that data is going to disappear. It might disappear at one site, but not everywhere, so

now you are worried more about physical connection which certainly is an issue or a non-kinetic, you know, disconnect, but that would be temporary and something that our cyber warriors would overcome.

It's the sword and shield problem. I do believe that our adversaries will adapt, and they will try to counter where they think we are going. We need to use some subterfuge and have multiple paths, to make their problem difficult. And then we are in a dynamic fight, and I believe that our agility as American warriors is going to be better than that of the potential adversaries that we'll face.

SPEAKER: One final question here in the front?

MR. CLARKE: Colin Clarke, Breaking Defense. I think this is a question for General Otto. I'm wondering what the balance is in your focus between efforts like open mission systems work, to make sure that data gets shared across platforms, and directly down to the war fighter. And this sort of deeper ability to have this enormous amount of data and share

it back and forth.

LT GENERAL OTTO: Well, we've got to have both. In my mind, open mission systems is an absolute imperative that will allow us to have the agility that we need going into the future, and then the access that we've been talking about of course, is how do you make sense of the data. So, there is the integrating -- I mean, it's really the five things. You've got the collection platform, you've got sensors, you've got data, data paths, you've got data storage and you've got the analyst.

The added capability comes from all five of those together, any one of which won't get us there, and so we are working on all five lines of effort in order to take the ball further down the field.

MR. CLARKE: But you are going to focus really far on one at any one time. That's what I was getting at.

LT GENERAL OTTO: Yeah. But I've got teams working on all of these. It's not the same people but, yeah.

SPEAKER: Well, gentlemen, I want to thank you very much. And I think we are all fortunate to have as much brainpower up here as were able to gather today, and it's been very insightful, and interesting to listen to. Please give our warm welcome for our folks. [Applause] That concludes this, don’t forget about visiting the AFA booth downstairs when you wander by.

* * * * *