# Transcript Day 2 video 2

# Session 4: Outcomes/Applications

### Break out session reports.

Next order of business is to have some report outs from various working groups

#### Defense and Intelligence

7:45

Andre

so defense and intelligence and so in the breakout group one thing that was clear we it's good to start with a good definition right students love it and professionals love it too because you can build on some solid foundations and so offered up was **data science: the art and craft of people making value out of technology and data** and that's it's pragmatic it's workable its suitably fuzzy and warm you know art and craft all right yeah we can go to the to the Saturday market with to the fair and but I think it captures well that it is about it is about technology right if you don't know anything about tech know you probably can - it assigns but if you don't know anything about the domain in which you're operating or about the people that will be operating in the domain including yourself then you can't do data science either and one of the most common complaints for example about data scientists coming out of academia is that they know how to code and throw data together but they don't know why to what end for what purpose especially in the business world but it's also probably true in the non business governmental world so that's **the art and craft of people making value out of technology and data**

so then we had four common themes

one was - maybe a need to **redefine we understand the concept of the map** what mapping means what artifacts would qualify as a map and what are some new challenges in that where altima t it comes down to the purpose whether we call it a map or a mapping effort the purpose is to get the **right information to the right person at the right moment in time** and so with that each of those right information right person right time has particular challenges and one thing we can see there is really that there is a conundrum of operating in multiple spaces simultaneously so we all operate obviously in this domain we think predominantly at first about Geographic space physical space but when thing and that's the space in which and we talked a lot about uncertainty how we convey it how we measure it how people can understand that how users can understand it but mostly we've gotten pretty good at doing that in the in **the physical space in geographic space positional accuracy it in other words that we can draw ellipses on a map** maybe with it's an ellipse right so has some variation two dimensions and and two then we pray that decision makers understand what that means though more often than not they might ask the analyst or who sits between the analyst and the decision-maker okay 50 percent chance 60 70 80 10 percent so so there's the question of how do we actually communicate that but even beyond that as I said we operate mostly in in uncertainty communication in the physical space domain however **when things go wrong it tends to be in other spaces namely attribute space** that for example the right building let's say is targeted and the and the missile is is the targeting is extremely precise it just happens to be the wrong type of building that was hit and that's that's an **error in attribute space** right when say and and an embassy is is hit of a foreign country as opposed to the intended target has happened in was a Belgrade right **but there are other spaces when things go wrong like if we miss understand or don't even know structures in knowledge space like for example ideological directions** that we don't even know about when people you know confuse say the written text of a particular religion is understood and highlighted from all kinds of ways but the oral tradition might not be as well understood which actually drives new ideologies and **so there's all kinds of spaces in which we have to organize them ourselves and then communicate to decision-makers** where we are and where we want to go in all of these different spaces **and it's not just physical space it's also attribute knowledge and network space**

then there was a warning a warning sign to step back from the technology zoom out and **understand the the application domains the the tradecraft and the workflows** and the question was asked how can we support that better whether it is training up domain experts in the technical side or getting more exposure to the technical experts with respect to tradecraft and domain knowledge and generally the way I would always refer to it is that it's **oiling the knowledge ecosystem of the geospatial domain** and the question is how well oil that ecosystem **is how easy it is it is it for example for an analyst at NGA to match a particular task against the set of hundreds of tools in say ArcGIS** because you cannot know know all of them but you have to quickly find the right tool so there's a lot of knowledge transfer needs from experienced operators that however are not up to speed on the data science approaches to a recent graduates who know tensorflow and that's their hammer for every nail so I think a big theme was over overall enabling the work for force I don't know that we had answers oh yes there were some answers as to so now I try to remember this was before lunch to me every day is BL and PL before lunch and post lunch so it was about small steps basically I would I would translate it as so if someone uses Excel now right there's nothing wrong with that it's important to build on that just to to say okay you have Excel okay how organize how well organized is that if you're looking at your existing Excel table that represents this situation does it actually make sense to you how would you improve it if you know really think about it and step back okay now beyond that the the example was was used okay once people get excited about even working with boring Excel tables they pick up somewhere I heard about this Python thing can I do something with that yes go for and marry the to marry what you already know with these new tools and and if it leads up to something useful being done using tensor flow so be it but if it's a simple regression so be it and perhaps it probably be the first choice so that's what I had

#### Insurance and Health

next we actually had an interesting thing happened on the way to the bay and the walk of the Insurance Group it became the insurance and health group in some ways so insurance first perhaps you think and go ahead and tell us what you two talked about

yes we just exchange ideas of the you know business you know ban this company my company is doing and I think later on I think we have people from the health as well they're doing research so we talked about the data you know how the you know **satellite imagery for example is helping the insurance industry you know a lot of for example the use case is underwriting** right so a lot of underwriters so they want to know that shoes up the it's probably casualty insurance they want to know you know what the property is like you know what is the **condition of the roof for example** and there are these you know **new generation insurance companies like keepo right so when you apply for insurance policy you know when you enter your address you know that policy form a lot of information will be pre-populated** and all these information is you know retrieved from somewhere but that's the data behind this thing we're know it's a lot of these imagery based assessments

and **another use case we insurance is in catastrophe could event response** so I think there have been talks about disaster management disaster analysis but then there's also you know for those kind of events there's the insurance perspective of it so when a hurricane you know came **in insurance companies would like to know where to allocate their claim adjusters** and they want to have a you know **rough sense of you know how much damage** or I have to put up with this in this event so depending on the path of the hurricane and wildfire is another example right the insurance companies want to know you know how much damage am i up to you you know can I sleep at night so those kind of questions so for disasters from an insurance perspective

so my background like my previous job was with a **catastrophe company** so I know like for the modeling piece as well for each peril like flood hurricane earthquake you know like there's a lot of geospatial application there especially right now **flood modeling** is very hot in the modeling world and that requires a very high level of accuracy in terms of cuz you know it matters a lot for just a few meters for the water damage so that's a you know another very another application area where geospatial data and you know the advancements in geospatial science you know has a big impact

so some other areas I think I'll just share some of my experience or what I heard in the insurance industry so **for traditional actuarial modeling for them to price insurance policies there's the spatial smoothing** just think about where your homeowners insurance or auto insurance if you're from neighboring zip code you **don't want to see your price to be dramatically different** so even though you know like running through some other you know GLM models and things like that you might get some big differential in the price but after that you know you can't just take the model output as is then you look at these kind of things you need to do some run through additional smoothing algorithm to make sure you know people can't you don't want people to take advantage of those gaps avoid ever selection and

so just finding general liking insurance industry geospatial applications it's just see a lot more on the uptrend people find more and more really helpful insights from location intelligence basically yeah so another I just saw about this so when you have a location right like for insurance you also want to know for earthquake risk for example you wanna know the soil type for hurricane you wanna know you know distance to coast those kind of things so **it's all very location-based** so with all the **information you know from the imagery and from you know modeling companies you know all these information gets enriched to that location and that will tell underwriting** a lot about you know what's about this risk

**health description** Jay from UC San Diego and we had a really good conversation it was a Georgia and Benjamin on the walk to the to the Bay and I think one field that is not being addressed so far in the location power to get rayon today is about the **crossover between geospatial technology data science and health** and what we're doing and what we're seeing a lot is is going to direction right now one direction is going to population level nation level international level **groups utilizing satellite imageries and sensor survey data trying to understand a neighborhood characteristic and the risk factors that directly impact certain health or certain disease outbreaks** as towards that direction there's a lot of work there from satellite imagery for developing or not really developing countries you don't have enough ground power or data collection so utilizing that you can understand for example we do see projects identifying the slumps different type of storms and in Mumbai and how to come up with a plan when there's a disease outbreak there's project that's the project going on **with supercomputing Center at UC San Diego** and

what my personal projects touch more on is utilizing patient level individual data with participants or **patients that we're sensors and we understand their daily exposure to environment what type of environment whether it's physical environment or nature environment including air quality noise** and how does that impact their health outcome and we're working on a lot of eating prediction so using **geo AI technology trying to understand the moment of eating** so with heart failure patients it's very important when they're eating when we can identify high propensity of eating we can suggest nearby healthier options so that's one of the power of ji-woo is location time in understanding the person's behavior so I think this there's a lot of uncertainty there but there's a lot of opportunity and this is what we're working on in precision medicine for you know trying to come up with customized individual treatment plan that works for everyone and there's specific condition I think helps. there's a lot of opportunity there are people start working on public health is already recognized impact on environmental health outcome and they're catching up the field and they need a lot of geospatial knowledge and technology help the fields so I was talking to George that we can get helps into one of the domain she sees the future direction maybe a lot of collaboration. Ajay also mentioned that some things work on that so yeah very good that's

very timely because next week we will Ajay will start the health domain working group so the use cases we spoke about are really very relevant to where we go with that groups its **population level and precision precision level**

24:43

#### Transportation

Reported by Ed Strocko

we had a robust discussion and a great group we quickly honed in on **HD maps** there was focus of our conversation there and we started the conversation about the difference between **machine learning and machines actually pulling data or generating the data with autonomous vehicle**s and how we need to think about that a little differently and that led into a conversation about the HD maps a mapping system we might need **to rethink what a mapping system is for these HD maps** and all the different layers in there so it's really a new approach we quickly realize that there's both technical and policy issues that need to be addressed and we really want to figure out how to make a standard or respect lasts over time and need to address some of the high end issues high order issues before we can get to some of the technical issues a lot of the questions **were what happens if there's a crash who's reliable is that the OEM is that the map maker is at the city and who owns the map** so a lot of discussion on that and how that fits together they're assuming that there is some type of entity that collects this and we develop standard

it was the recognition that the this space is different than some other transportation spaces like aviation where airlines don't compete on safety where you know in the Surface Transportation based people do compete on safety and so **how do you protect that IP while still giving some of the base data to an open common repository** and so really thinking about what that what that looks like how that gets done how that gets done in a production environment and don't have copies of this all over the place the structure of that and then

we heard a couple different examples we talked about Singapore and what goes on there and then right here locally **LA with their shared mobility standard** and what's happening there and very exciting approach there and how they've not only established the standard but also a nonprofit to deal with some of the issues around that so we're thinking about what an MVP might look like for this and you know what our next steps could be in this conversation so recognizing this is different types of this different type of data different approach I really want to look for that that consistency between the providers of the data as we go forward and then hook up with the Geo community a little bit more probably do a gap analysis and really play off of what we're seeing in LA as a starting point for a standard and build out from their (Get notes from Ed Strocko)

28:20

#### natural resources and agriculture

Regan Smyth reporting

we talked a lot I'm trying to remember now but we talked a lot about one trust and what does it take to make sure the analysis the analysis you're doing get uptake by folks and what some of the issues are there you know **the idea of going in as a data scientist with the answers we the kind of collective experience was often that was really counterproductive** and a good way to make sure that nothing you do serves any value in the world because people come with their own perspectives and they want they want to understand what you're doing and they want their voice to be heard and so that's that's a important component of the process

along those lines too we explored that issue that we started exploring as a group of that human component and how do you build that into processes what kind of information you get and what situations do modeling approaches work and how do you know when they're not and are you able to address that so that that was another topic of conversation

the bigger challenges and you know doing **running various models but not necessarily understanding what the drivers are and whether data science can help you identify those** or if it can kind of mislead you in different directions if you you're not measuring the right things and learned some really interesting stuff about how things are measured on agricultural lands and some of those data sources

we also had a conversation about **data sharing** and kind of what what's out there being collected and how people do or don't have access. What scale we should be operating at and how important scale is or isn't you know is it always the best to try to get the most precise information and make inference based on that or is that both unnecessary and inefficient and sometimes gives you the wrong types of answers that was conversation

what else did I miss anyone who was see the people in my group all right well that's some of it at least

#### Data Sharing and Ethics

31:13

Jeremy Morley

we had a great walk okay yes so I get to look to some of the people who with me because we inevitably I think strung out into two or three groups or may not have got everything here

so we discussed kind of data sharing and ethics as we as we went in different parts there was a **general feeling that of course data sharing to be encouraged** that maybe on the kind of EO kind of observational mass collection side there is possibly some thinking that it's **the difficulty there is in actually acquiring things likes a ground truth or reference data** that goes with you know and how to encourage sharing of that and particularly there's a point about there's being a history from agencies like ESA funding projects over years and actually trying to dig back and **find previous say ground truth sets from previous studies is not so easy and what what might be done about say repositories** of that sort of information

there were different certainly different models amongst the groups with us about how they data sharing is supported and paid for cuz put data out there and **to support it sharing is not a cost-free activity** per se and then

in some in some groups there's thinking about how to build in d**ata ethics** in at the start with the data sharing and and making decisions a bit more particularly about what might **be personally revealing data** that's being shared and how to handle an or **anonymize** that's that sort of thing an interesting discussion evolved around **what is actually personally identifiable information** and how you identify that and for example these days if you are starting from scratch with their telephone systems would you be able to publish a phone book anymore with the list of names and the names addresses and phone numbers and all the rest

and then also aspects of bias as well both on the sharing and the ethics side there about either handling it an intentional data collection and going back to some things that came out from the panel speakers of **understanding bias** in the in the data and how you how you assess that particularly where you don't know the whole population so it's difficult to know whether your sample is biased in that case but also on a flip side of there are some sources that just will be inherently biased like if you're trying to extract stuff from Twitter yeah there is no not biasing sort of way of sampling from that so **then it's a more of a matter of understanding what the bias is not asking the right questions** from that and being ethical about actually kind of asking the right questions and also in that space understanding what you have permission to do really with that that data not in the legal sense but in the ethical sense of **what are people expect him to be done with that that information as part of that that ethics** as well so I'm not sure we came to any grand conclusions about how to solve those things but at least he's described a bit of a problem space

any others who were on that walk if they missed any topics in in that takes answer yes it was **perfect timing it was 20 minutes out to the end we saw the vista and 20 minutes back**

# Session 5: Actions to Take

Chair Nadine Alameh

while we're waiting I got my laptop my sticker they're brand new I'm so proud

38:09

this session is all about actions so essentially what do we each one of us take out of these last two days and go back to our real lives and what do we do differently what do we you know try to interject into our daily jobs out of these two days

so a couple of things to keep in mind one is you know actually **Adam set up this page with which allows you to post questions that the panelists and the keynote speaker**s can answer you can even vote so we have actually some questions up there so if you like a question just vote on it so because there are many and I think this would be one of the leftovers of this workshop of this summit we're not going to be able to answer all the questions so this is at least would give us the priorities of the people in the room so please you know check it out and vote I think the panelists would appreciate it

and the other thing before I introduce the two presentations as again this is about actions so as you listen to the presentations and to our panelists think about

* what we can do to bridge data science and traditional GIS you know both ways
* what skill sets will be required how do we train people how do we train data scientists
* how do we help organizations automate and scale geospatial data so that it is you know usable in workflows that involve artificial intelligence and machine learning and data science
* what actions could a GC do to help with this and any other suggestions for the future

so keep it's a lot of questions keep those in mind as you listen so with that in mind

## Presentation on Actions Satoshi Sekiguchi, AIST

40:30

our first speaker on the actions again so we're going to start doing things here as dr. Satoshi Sekiguchi and he is the VP of the National Institute of Advanced industrial science and technology or AIST in Japan and his team is constructing the **world's largest AI focused open public computing infrastructure and they call it ABCI** and he is going to tell us all about it's all yours

42:19

okay good afternoon thank you for having me in this very exciting now workshop today in particular for George. In the listening to the discussion from yesterday and today so the I'm feeling a little bit nervous about that because today I'm a right person to speak about in this kind of on the topics there is because. I haven't been well yeah I used to attended the OGC meeting and around ten years ago but yeah last ten years sorry I'm a little bit here staying a little bit far from the OGC activity so you're more like doing the management stuff rather than doing the research so that's why

my talk there would be something back this way

so the as I mentioned he what ten years ago the OGC at the ast the my team they started a so-called **geogrid** **project** so which is the intended kill offer well they actually the at that time the ER batil four five so the the even the Google Earth as not being in a public space so that the the we used to have in a I see we have a we have a satellige the imagery data set there from the ASTER sensors but yet those data's archived on the tape so that if you need to access some part of the such an allocation data or imagery data whether you have to wait at least a one week to get the such an image from the archives so okay so the I or my team me coming from the IT side me sitting in between the satellite imagery data group and then the application group so that we can offer more advanced and IT technology to bridges a gap so or the first thing we did was to archive to bring all the data into the disk space so that if we can access such an edit instantly so there was the original idea of the Geo grid and then **they're still doing lot of the application on top of doesn't our infrastructure** so the actually at that time the AIST or the **open grid forum** we called as sponsor member and then as well as founding members and also the member of OCC associate member since the Year to us so that we used to work very closely so the briefly mentioned that in the aster is the

at that time actually the 10 years ago 12 years ago there was our main contents of the our geo grid so it's an optical sensors and then the covering most of the landspace and then he provided a very good the maps and they out of such in the sense of data and are still be active actually the nominal a lifetime. the satellite and launched in year 2000 and then we still working and very healthy and then the for example using the such as the Aster Data so which is the the flying over the earth and then the revisit the period is about around two weeks so that the we taken the in a picture and to find out and the kind **of time series of the changes** of the same location so yeah for example actually this is the case of the Bangkok new airport so the taking the pictures since the year 2000 till the 2006 so that we can find the great changes how such an a big Airport and then the we can find an a similar other changes so many places and then the also the e we not only we yes and the move the data from the tape to disc but also early we are thinking there how to you given a services to for the for the users but the well **we decided to use the OGC standard services** one to capture the part of the year such an the data and then also if possible so such an the data can be processed in the remote computing services. actually even at that time the not only just in the handling such an imagery data but also the we're thinking how to deal with so huge amount of data in by using the large scale of the computer system and

this in the one of the **example of the workflow** all of the planning of the **mitigation from a disaster** well the top left so they're like the Geo dead geology data which also a part of the AIST mission but the solving the geology in a country and they also uses a satellite imagery data from the ASTOR the me create a kind of digital elevation model and then again and also the in-situ data it's all combined together and then the using the big computing system and that is **simulate the sort of in the lava flow or some other disaster** well as short as possible and then creating in a kind of an ahead of map and also this is a case of the landslide right so the again that taking the geological data as well as the meteorology taters and then he also the data elevation model and then creating the **landslide a the simulation model** and then they gather and all data and then to predict or we happen in the next one week or something so that the you know we had a lot of the study even twelve years ago

then the **identified the sort of the architecture** for making happen so I would come back on this slide the later but this structure the it's still viable I mean the from the bottom so there are the resource of a year and then the for the computing facility and data storage facility and then also day we can access to the Internet data's and also for the managing such and data we have to consider the security issues and how to communicate such data to each other and then the how to give the authentication and authorization of each part of the data and then this there are similarly layers from the system layers up to the application layers so they were a similar the opportunity to for the having either kind of the **platform for offering the data so that data access service or what workflow engine services** and etc so that's kind of the that template you know for all our geo standard services so

they a little bit the different topics I'm going to talk about the which is the in Japan the government is still government is strongly pushing on artificial intelligence and its R&D the in the past couple of years the **Japanese government is strongly promoting the Society 500** and then the which they even the scientists investigating the year so I five dot all so well this is actually well we kind of **digitizing the whole society** so that not only making a application in each silo but they were we wanted to set up the sort of the platform where we create the application on top of it so that the making the s**ociety 500 happened they of course artificial intelligence is key technologies** so the but the indeed the AI so

you might you know family with that yeah we used to have the **artificial intelligence boom several times** and then the first one they actually back in 1960s or some time around but a second big wave was actually in the middle of the eighties so at that time fifth generation computing the project which was initiated by the Japanese Ministry meeting and then the even at that time the Japanese the historically in our approach to AI he has always included the infrastructure and the software and also this kind of the concept leads to you our current effort within the ABCI and then well the third generation comes in 2010 time frame so that we were more like focusing on the deep learning types of the applications and then the the government is stating that the AI for changing the industry of course they are there are a lot of the such in the statement around the world

but in a Japanese case so **small medium enterprises** okay so they were very interested in you introduced such an artificial intelligence technology to change their business model but they don't know the how to do it and then again the left path the only 8% of the small medium enterprises are actively using the artificial intelligence technologies so this Friday they actually they study in 2017 but well the percentage of certainly active portion of the small medium enterprises doesn't change so this means they are still the **80 or 90 percent then considering how to use it but they didn't know what in the best way to do** so so the government is strongly pushing and but the SEO is feeling very very serious headache and the they have no funding or the investment to make the artificially intelligent me hack

so the in in this circumstances AIST is my organization and part of is METI and the Ministry of Economy Trade and the industry and then the we deal rather apply research in the basic sciences and then the we're the cross walking together within the industry well historically the my or our original the laboratory started back in 1891 but Wow the being a quite long history of our reseach institute. but again the **we are sitting in somewhere between the academia and then the industry** sothat the our mission the trying **to bridge the gap between technology and their business** so what this situation is the somehow different from the other part of the world they were happening but that **we're listening the what the such an a private sector wanted to do and then okay so let's think about the let's find out the solution to fill the gap** so

**under the AIST there are seven research department periods** and my department is in the information technology and the human factors and then they also the right table you can see that there are metrology Institute of Japan it's a close and a function of NIST here and also at the bottom is a Geological Survey of Japan as in a USGS so that the IT and then the standards and then the geo data is all under the AIST umbrella

57:00

in my department the I started the at started a research center around five years ago and it was somehow the very early a Research Center to serve as a **core hub for artificial intelligence research** we invite **Dr. Tsujii Junichi as an ahead of our research center** and then the we have doing a lot of the research regarding the mobility of manufacturing industry etc

so given the situation again the way across three well we offered the Research Center and as the core and the I and the **data science people the coming together to our the research center** about the and then the they also be their big data's that we can make captures and then we can access from the internet and also we have an important shape with the industry so that we have a lot of the opportunity to have in a big date of course the set satellite image data is another the interpreter and then the deep learning D planning is a new engine which means the you can you can do the main things out of such an acuity but here for making and then happened hi performance computing it and a key the component to make the real big data and then the deep running

so the status I mean the traditional **HPC application** so that there isn't more like the you can have a **mathematical or physical model** of the here and then the you can create a new computational model out of such an authority then the if you if you need a more than fine granularity a finer granularity in the you need more computing power to make the more precise and the result out of such an the simulation but there are a lot of the opportunity and then there is the space for the traditional HPC system the creating an a value now but

on the other hand that the models can't be mathematically formalized but the this type of the complicated or complex in the situation like and human the people's mobilities or the understanding the real world and then the that kind of things know were formalized by the real mathematical formulation so that we're making this kind of model by gathering the Rada with the data's and then the processing such an the data and then create on the computational model then the if you give some I this situation than the predict in the future of the system so the this type of the ideal world or no problem yeah we need to gather or big data area

so the algorithm computing in a big data is an a key component for the current the artificial intelligence but still missing pieces in computing. That is because the cloud service provider so but they don't see the promising that customers so that they are very afraid of the putting a lot of money to prepare a large scale of the computing facility and then the also on the other hand the client ok so we need to do something but yeah we have no space to try to run our application so that this is a kind of a **chicken and egg problem** so that as we are **sitting in between the industrial sector and the academia** okay

so AIST and happy to offer you the kind of the cloud infrastructure to one to help there the big demand and then the what the so that would my world I spoke with the head of ministry and then the asking them to have to secure some project the funding to develop the system and which is right now the we achieved the hundred fifty AI xFlops. x flops means the using the sixteen the floating-point the representation so that what that mean half precision of the system and then the if we use the double precision we achieve with **thirty seven point two petaflop** so there is a one of the highest the score in in Japan right now the **top 500 list** and there will be renewed maybe next week but this is a previous one the latest one but the **ABCI spotted on the eighth ranking as a high speed** and also the their **energy saving** in the green five hundred their performance measure is spotted on the third ranking in the world so

the unique feature of this in ABCA its the offering the large scale of astray services there which is compatible with the Amazon s3 competing Amazon s3 the api's so that the try our client can access to our system the something like the Amazon s3 type of the interfaces and then the also the we have the hooked up with the high academic a high-speed network which is about sine of five so with a hundred at least and 100 Giga access to our the cloud services other than the people are getting through this and assign at services and cue the ABC are quite a **quite high bandwidth** and then the the of course the we were working together with the industrial partners so that we secure quite high level of the **data encryption** or other security services and for the **software stack** do this as standard the software's but the **most of the internal software comes from the HPC operational and the tools but somehow combining that widow of HPC made the AI components** **like container engine** so that kind of the hybrid system we were the offering and then

**for energy saving systems** of the year maybe I have much time to talk about it but we implement the hybrid cooling system so that we don't need well even in some hot summer in Tokyo area but we don't need to operate the chiller until creator the cool waters cold waters they very hot part of the this computing not easily NVIDIA I heard yesterday in visual person talking about it but yeah we have the 100 quad we hunt 300 system at the single and the these are very hot part so that we use in a kind of liquid cooling and he'll take the excessive heat out and then also the other part is using the standard air for but still we can give the 30 kill siliceous waters into the system and that taken out day of the lower than 40 the associate's degrees so that maintaining the eight associate degree that the differences isn't a key if you make this system heretic so the even and hot summer day we year survived without using the chiller and I officiate intended kill in the give this system for the several layers of the people the from the expert advanced intermediate and then the novices and a for the export the we were measuring the the benchmark and then the couple of years ago

so there are no much the people are they participating in **this MLPerf benchmark but after ABCI got the top score** of this the benchmark so right after a week the Google with an CPU you trying to beat us and then also it is still doing their lot of the very hot the competition and then right now the **ABCI have a team and then the Fujitsu is the number one** holding a number one scores and then therefore the other the purpose of this and the ABC a system the we're we have the playing that GO system on top of this the ABCI so but it is still challenge and the father the

1:08

last button by next actions so ABCI kind of a harsher intelligence file meaning that the even such a small medium enterprise they don't have the any of the facility but the ABCI is an avenue where such an small medium enterprise to ask and then to create the this kind of an a component okay officially happy to make it and then back to you the SME so this is like semiconductors they're fabulous and the fab pipelines so

we are offering the **small medium enterprises so such and very simple or they start using deep learning** today we know **Sony's neural network console (Vide0)** and also create a new project and a lot of training for the image signals and any type of data and presented as vectors and matrices can be loaded compose functions instantly with a click of a mouse drag and drop copy and paste to quickly create the large networks finish to design a network it's time to Train it **powered by our neural network libraries** in your network console trains your network fast and originally this system has the develop after training evaluate the performance of the network we develop the interface compared with other networks you design data grants the training history of each networks is saved for the later browsing is the automatic structured search future to optimize neural networks the neural network console automatically **searched for the lightweight high-performance neural network structure** for you a rich amount of sample project will help you on the way to quick master deep learning [Music]

even a novice doesn't know much about the deep running of the how to build a network but and then even the system is running in a desktop system also so you can just click on the server the system so the network's they can be very simply run on our ABCI so you can scale the your network quite simply and quickly and then the some other an extra action I'll be a little bit a touch on that and then the and there may be a joy which in the Kim-sen it

**Kim-sen is nurturing in the OGC GeoAI working group** and then they which is trying to make the space in the information more personalized so what that means so the you didn't need to well the we have the simple add a single on the platform where they all hope so the data is gathered but the in the context of the each of the human you can filter out some of the information but here for for make such an things they happen we need a very large scalable computing

1:13:13 Video Space Personalization

we gathered in the standard format but depending on the context and the filtering out of a lot of the information and then the more the information how to create a **3d data generatio**n out of the similar situation and then LIDAR information is priced into the servers and which can be braked out into the several data servers and then the other one such error and the point of a crowd data is dividing in theory several sectors and then the Earth's out there gathering such an i the the daters and then the the users object models and then to **create the real 3d model** and then the such a new space object as worried that the small object like on a chair the table etc so there can be used the **uses of the AI technology** to the recognized have no or is what sure that it's a similar thing and

finally they were still the working on a **satellite data archive in the AIST** in the yesterday the I'm hearing that yeah after the flooding at in a tropical area so they are run over the cloud covers the area so the optical the sensor can capture the images on the ground but for the SAR data the you can see that the ground and also the horizon the data together together with an we can a more lively the image can be created but the original saw data is actually black and white and so that the we are the combined together with optical data and **the SAR data and by using a lot of the Deep Learning type of the application** that you see there what an earlier color the under the cloud so making the training in the such an SAR data with an optical data and then they make there's that kind of the image teachers out of using the combining the using of the studied and then the optical data's and the industry they have their

1:17

future actions the again the we have to consider the modern more modern architecture but the skill the **layered architecture** is the same and but they may be in the middle space I am saying that there is no open we're in the service but **the model has been they change it by saying that your cloud service** or et cetera and then also the way of planning into getting into the ABCI into the all up free and then we can accommodate this kind of that services

so and then the even in Japan where they we have **commercialized and cloud service and the providers** vary within the chakra which is offering the terrace as an open and free such energy Wow shattered imagery data and geo date and a platform so supported by METI so this is the kind of the great challenge to make the such an a data for for the business tell that so-called open freedom platform and and be well there are of the internet foxes and the working internet service people are trying to make an a business on top of this and the **Tellus** services and of course and the standardization is in a key for the future

for the summary the again the ABCI has been operational since the to the last year they're supporting SMEs and their area challenges and that geoAI is I'm promising an application on top of ABCI and then other high-performance AI and then quality of a satellite imagery is getting much richer combining optical the data and circuit which is also powered by the ABCI and their **future actions proposed I'm proposing the including in a well again the architecture standards** and then the back to the such an movie but in the easy access to you such an the monster power so thank you again anything for your kind attention

[Applause]

Question: any of these systems open for international researchers?

Satoshi-san: yes yes. we need a contract of course but it's open up opportunity for everybody

Question: ASTER data free from Japan?

Satoshi-san: yes and some of the but you can't download it and so then you have to work on in the only tell us

Nadine: my favorite part was the standards action thank you

## Presentation on Actions: Andrew Brooks, NGA - Title: Designing the Future of Data Science

Nadine: so first I wanted to present him as dr. Andrew Brooks but then he said he spent enough time with us these last two days we can call him Andy so I'd like you to welcome Andy he's the chief data scientist NGA his LinkedIn page is one of my favorites because he says his **job** **title chief data scientist what do you do one line create value out of data** and that's what he's gonna talk to us about

1:21:17

so the title for there would like where that came from is when I joined the IC because I've been there about two years so and I've been in the industry before that you know in here in the Bay Area on the west coast and it's one of like oh you're going into the IC and it was truly like when I got recruited from my job it was an unlisted phone number that you know you never answer one of those like one your phone rings and I was in my kitchen I was about ready to go out for a run and it answered and they're like hi this is doubted enough from the intelligence community and I stand in my kitchen in San Francisco and I looked out the window and I'm like what is going on like I thought this was a prank for about 20 minutes or so but truly that's my job I have kind of a rare role in the intelligence community particularly with NGA men on public and what I do and a lot of the work I talk about is public and what I do **as the chief data scientist is essentially what I do is figure out how to create value out of our data** I've done that for about 20 years or so all with tech companies and startups corporate data brokers Yahoo research things like that out here. there's some new form of data what are we gonna do with that right and it's I've done that with supply chain data health bioinformatics my dissertation was on physiological data heart rate variability if you don't have that specificity things like that but this is what I do is figure out how you create value out of data whatever that's going to look like and

what I want to talk to you about today is the efforts that we've got at NGA to be able to do this how do you take NGA and I'll go into an overview of who we are what we do how do you do this like it's scale right how do you do these sorts of things as I jokingly say **we can't tensorflow our way out of this problem** or put things in the cloud and things like that this is not a technical problem or a technological problem this is bigger than that it's more like a cultural people and data sort of thing so I'll go into that today this **whole talk is unclassified** everything here you can ask me any sort of questions like that I want to try and answer as much as possible so we'll save some time for that if we do bump into things where I can't disclose that I was call it outright you know and just say hey I can't answer that right here but that doesn't mean I'll punt I'll take that question and go find out see if I can get an answer and communicate it back because I want to be able to answer questions from the public and what we do

so let's talk about this almost like to give me obvious slide **like where are we today with like data science** but you can do a geospatial let me pause there and go is like well what is data science and you heard it earlier it's I **described it as the art and craft of people leveraging technology to create value out of data** right that's the way I look at it I'm really pushing that within our agency but more broadly and what we do and that this is a real like you can say revolution that's happening or is changing and such and **that it's more than just big data sphere it's more than put your data in the cloud it's more than use NLP it's more than using AI this is like a top-to-bottom rethinking of like how we do stuff how we do our work** and that's what I want to go into today with NGA and what's happening is just like the obvious right so we've got **more and more data coming from more and more sensors more and more techniques to be able to do things for us** like in our space like a national security it's like we go like pretty calm potent let's say adversaries Russia China the two plus three the way that we describe it have some incredible capabilities in this space and are moving very very quickly which is a big focusing mechanism for our efforts in our agency where do you start with using data science right you could cover so many sorts of things and one is rallying towards focusing how do we best counter and protect national security for the country for what we're trying to do

so like here's where the world is really today and what I like to pushing in into next is like where do we think it's gonna go next like what is it are we just gonna get more data are we just gonna get more sensors it's gonna get more complex and those sorts of things and **as an economist I love to try and predict things I'm a prediction market kind of guy but like I don't know where this is going** to go right like what the framework is or like how the future is going to go and placing bets on that and that's not really the space that I'm in and particularly is like in national security the Intel community is placing bets like that we just don't do it **what I look for instead is like okay like what do we want that future to look like and how do we design it** and that's much more of what I'm going to talk about here is one where how do we take efforts in are like individual work roles so like specifically like **how you spend your time as well as like on teams and in organizations and particularly organizations** like this how do we collectively move towards this future that we really want and I kind of I'll project here a little bit on what what we want that future to look like one where it's like like geospatial data or data in general it's like for good right like let's just say that like we want that to be for good **we want data to be collected analyzed and provide value to users to people things like that move things** forward right and I want to say that just because it's like that's not always the case like there's a lot of instances we'll read about in the press where data is collected particularly geospatial longitude/latitude timestamp and is not necessarily used like a really positive ways replicating what has been done let's say example with redlining if you're familiar with that and doing web lining and all sorts of things and kind of going these like break break that's not what we want **we want geospatial data in particular to be a positive force for good and knowing when we say that like things get awkward like that's kind of hard cuz we got to figure out what good means** and how to do that and that's what we'll talk about in here one of the things I look at like in my work is like just kind of laying that out there like

1:27

where do you begin we're doing efforts in this space right how do you look at it one that I like to **look go and try and find it's like existing frameworks** I'll call it or models not really like in the data science model kind of sense but like ways to help kind of guide our thinking one that I've pulled on lately from a couple weeks ago was from **Stanford's D school** is anybody familiar with Stanford's D school up 101 here awesome ok great so this is a slide that they put out a little while ago and I'll zoom in and talk a little bit more about it but Stanford's D schools is really awesome and that's an **interdisciplinary program** you can call multidisciplinary program well what is that it's a program where it's like the people who go there are not like designers they're like regular people right you know and they're coming from different disciplines and backgrounds and things like that and they come together to try and **solve like complex problems right so you'll get like a computer scientist an engineer a philosopher a linguistics major and a business school student right trying to solve some sort of problems** based like in the world and things like that and what I love it is it has a very holistic look at like how we solve problems in the world very diverse stakeholders diverse viewpoints different sorts of skills and backgrounds focused on a specific problem set and trying to move in that direction of what they're trying to do now the D-school has been around for a number of years I remember when I was first inkling I'm like oh maybe should I quit my industry job and go into graduate school is hanging out at the D-school and being like this is kind of the way that I think right but it's I didn't go to the D school I went to Cal and it's like adopted many of these sorts of models and the work that I do what's fascinating is

I'll step through this and hopefully folks can read it it's a little challenging but it's one where they really look at it's like if we're gonna go out there and solve problems particularly with data and oh gee where do we kind of sit and this is not necessarily a ladder so to speak but like where do we sit or where are you situated and more so **what are the types of questions and things that we need to wrestle with if we're trying to create products services technology or the future of something** right you know what are those sorts of things and I'll step through this but it's all the way from like in this world is like the data right like **what's the underlying data** that we're working with in our world were like awesome we've got some new synthetic aperture radar data coming on what can we you know would is that what can we do with it right well that's like that's **one slice of this right what are the other technologies** that we're using on which the previous speaker spoke about it's like what are we using to say crunch analyze that sort of data other folks who are interested near particularly like **ESRI and others it's like well what are the products that we would create out of that** one that I'm really keen on and focus and we'll dive into is words **the overall experience for users like for us our analysts in the agency which I'll go into all the way down to door kickers would call them war fighters things like that what is that experience how do you consider that** what are the **broader systems** that we interact with right so like a system so like an organization or a culture or a place how do you do these things as well as then looking at **what are the implications of this type of work** and what we're trying to do what

I love to **use is this as a framework is to kind of help situate teams** who are creating and building designing things particularly in the space that are moving very quickly to kind of structure their thoughts like where do we sit and I'm gonna take you through a team in NGA who is doing this type of work is really trying to get folks where do we sit here what do we think about these other things as we're building this what do we need to be mindful of who should go hang out and talk to and listen to and learn from who may be doing other work in this space so we can then better create they design what this future space is that we want to do so I'm gonna go step through this here in a little bit

as a level set just as like **who is NGA** well this is from our webpage and we deliver world-class geospatial intelligence that provides a decision advantage to policymakers warfighters intelligence professionals and first responders right that's our official boilerplate of what we do inside of there the way I always like to describe it is you know **hurricane disaster response where the folks who can help figure out if that runway is long enough and if it's open to land a plane on it that's got enough water to help the people who are there** like that sort of flow all the way down to the decision-making part about well we've been tracking this guy we think he might be the **head of Isis** is this the place where he lives is this the map is this location is he going to be there what are the collateral damage if we send a team in there in a helicopter to go get him in a foreign land all that Kosta those are the types of things that we really work on I really like to try and situate that in the decision-making of like world space of our users and our customers and what they do inside the agency right so it's a big range of things all kind of like high stakes you know not like more or less important or other sorts of things but it's like it **kind of gotta be right** you know if you're using data science to work in these sorts of spaces like they kind of gotta be right because if you're not you're gonna end up in the news people are gonna die like the **stakes are pretty high** it is I like to say it's like we're not just putting little blue dots on the map you know like this is pretty high stakes type of work of what we do engi in general is pretty quiet you know most folks don't really know who we are inside the intelligence community like when I joined folks are like why all CIA I know the NSA and those other folks so who are you guys and I'm like oh **the map people** you know you can kind of just pass under that and then you can kind of just duck and just walk away but it's like that's not just the map people we do all this other crazy kind of cool stuff too but again like I said is being doing to the IC based out here on the west coast is you kind of be a little quiet about that but it's like we have a beautiful story to tell like what we do is pretty fascinating and amazing so it's a really fun cool place to be all right

1:32

so what do I do now is spin into a very specific instance of a team who is out there I would say creating the future of data science within NGA and the role that they're doing and the actions that they're taking and I'm gonna use the framework from Stanford's D school to help make that like applicable so it can grok so hopefully there'll be some examples there and then we can go into some questions on that the team that I'm going to talk about is one called it's called the **NGA Data corps** I would show you their pictures right but it's the intelligence community's you know only put people's faces up so it's like they don't even wants to put pictures of buildings and things like that well this one I pull from Trevor Paglen who's I don't know if Elks are familiar with him I love Trevor's art it's totally great it's cool it's fun and our public release people are always a little scared when I put his kind of stuff up but it's like this is our building so this is where we are in Springfield Virginia as one of our locations we have a tremendous number of analysts here we also have some in st. Louis Missouri as well as distributed all over the world where people are making big key decisions and such the team data corps I want to talk to you about is one that's just about two years old my first year I was at NGA **I was the founder of NGA Data corps and led that team and grew it from I think we had four of us to about 80 in a period of a year** and the challenge of that team or what they were set out to do is literally create this future of using data science within the agency because we had folks who are coming to us industry partners and others saying like hey put your stuff in the cloud hey use data science and all these sorts of things and there was a very great deliberate effort to kind of be like pause timeout like what are we trying to do like how are we going to do this like effectively it was amazing is the agency allocated a huge number of resources in people to go out and figure out what this is **so as founding this team it was one of how do we bring about this cultural shift in change and I kind of took this like social contagion Network Theory model of like let's get this team together of highly capable people with very diverse backgrounds to go out and solve problem**s with inside the agency so earlier there was the talk about like Oh Excel spreadsheets and things like that that was this team so going out and working with our analysts in the building who are looking and collecting and analyzing data informed decision making to warfighters and other to help them solve their data problems like if we want to figure out what this future is going to be go sit with the folks and go create it and go build it and that's this team of data corps folks we hired about 80 in a year the original thought was like oh you're doing data science so hire 80 data scientists and I was like no no **we don't want 80 data scientists** I need a chunk of those folks right who can do that where it's pure like writing algorithms and deploying them and seeing how you do all kind of stuff but much more I was like for those of us we'll all done this work like that's part of it but like most of it is all the like medical skunk kind of stuff right or it's like **getting the data prepared building pipelines** and all those sorts of things so split the model of the team up and such that **we have data managers data engineers data stewards data analysts who all like get the data ready to go get it prepped** and things like that keep it as many I say if not all of these folks have at least some background and like user experience research we and pulled that from us digital service in 18m like my prior work of getting folks who are like have deep technical knowledge and things and at the same time can sit down with an analyst who doesn't know anything about Python doesn't know anything about any of these technologies or something and can listen and empathize and understand what their life is like and how to do these sorts of things so this **team of 80 folks what they do is they go out and deploying in teams of four or five and go sit with our analyst users** throughout the building to figure out how they do their work right how do we solve their problems how do we create the future of data science at a very grassroots low level of like sitting with folks and being like okay how do you do your work how do we improve it what's this going to look like right so very much from the bottom up to be clear one of the things is **this is not opposed to top-down right like this is coming we're doing it in multiple directions but this is one that's very much situated in engaging with folks and pushing these things forward**

1:36

so what I want to do is step through this team and really how they operates like in a specific case with this model and how they're creating the future of data science in NGA and our place using this sort of framework and what one thing I really want folks to think about it's like what we're doing and how we're doing it it's not really just because it's like in the DoD or in the federal government or it's in the intelligence community **we are an organization like many others who have been collecting volumes of data for a very long period** of time have been creating value out of it in different sorts of ways have all sorts of bureaucracies and weird things and operate in the market or an economy and such and it's one where it's like the lessons from this team can apply in a lot of different places not just like in the IC or the DoD or the federal government so I mentioned with that team they're kind of a different background or experience and the work that they do and that many of them have like this user experience research sort of background so using the framework here from the d.school it's like where would they really start right a lot of folks what kind of things like Oh data scientists where they're going to start let's start engaging with users and talking about the data and it's like well already kind of gave it away but no it's like they talked about the **experience part** right so someone will come to them and say like hey I've got a data problem or I got like a challenge here's and I'm trying to do and they're like okay no listen to that right they're like okay well like we're gonna come on over we're gonna hang out and see what you do like what are you trying to do like what is that work and they always try **and situate at this team and really trying to fundamentally under the stand the experience of a user** now that sounds really abstract so make all make it more specific we've got analysts who's like job responsibilities is **to figure out like what's driving around in the back of white pickup trucks in West Africa** right like that sounds really nuts but it's just like we do we have people who specialize in these sorts of things so the key thing is going and figure out what else they're experienced in doing their work and doing their trade and doing their craft what does that look like right to get some real ground-level understanding of what it's like to be an analyst and this is really important because I forgot to mention you know these eighty folks that we hired to come join this team I think over **two-thirds to maybe three quarters were from outside the intelligence community** they were from outside the federal government they were from different places where it's like thoroughly know what I mean being an analyst means right which is kind of great like you're naive you can just go in there start asking all these like obvious ridiculous questions which is great but this is where they start they situate their work so really is going in and **understandings like how do you do this today what is the role of data in this experience what works doesn't work what kind of policies or are you running up against to try and do your work** stuff like that so trying to get that ground level very visceral learned experience from the folks who are doing this work

1:38

**from there they go down to okay let's start talking about the data** right where is this data coming from how are you using it is it **big data air quotes is it small data living in spreadsheets** things like that trying to get that sense of what that work is and how they do it right so going from that experience to going to let's look at the ground data part **we're purposely not looking at technologies we're purposely not looking at products** or anything like that just that fundamental what are you trying to do and like what's the data that you're working with so a lot of this really at this point is engaged in conversation and a whole lot of drawing stuff on boards right to try and really **map and diagram what that experience** is really like for these folks right which you can run up against people who don't want to do that because they want to talk about technology or products and things like that and this really exposes how people do their work and their craft so you bring people together to figure out well how do you really do it maybe document all this to get smarter about it right

**next phase that they go into is moving from the data part up to the technologies** part right so like okay I understand I got a sense of how you do your work and what it is the role of data in that so like get a sense of like where is the technology in what you're doing right what do they think technology is how is it used again we're not using products yet well go up the ladder and get there in a second but it's much more of like what is that fundamental technology underneath there and for our users in particular we're trying to work on developing a solution or a problem with them **what's their literacy with using different forms of technolog**y where are they with an incredibly diverse workforce people from all over people all different backgrounds educational skill backgrounds and things like that and the key folks with this team is going out meeting people where they are right so if you've got someone who's describing a problem and they're working in Excel if you start bombing in that you're gonna start doing stuff in Python and all this crazy stuff that's not gonna map like they're just it's not going to get picked up it's not going to work so in this is looking at what are the technologies that they have today that they have literally like on their desktop with ready access which if you and in the DoD is like don't have to be a ato whitelisted or anything crazy like that but like truly that and then solving that problem with what they've got there as well as I noted like the diagram with the data flows and everything mapping where this could go into the future and working with those folks and I'll talk more about that so the technology side and then you can kind of see where we're going and give it away is the

the product side right so like how does the experience how does the data how do the technologies and then **how do the products all roll together in that experience** of what they're trying to do so going to that um oftentimes we can find and give the story away is like really a lot of the challenges are running into have nothing really to do with like the product they'll maybe complain about the blue button doesn't work or something like that but really it's a fundamental things like the data is not really that good or the underlying technology doesn't work or the policy isn't enabling them to do their things so that's why this is kind of last like and what the team will look like is the product side time-wise

like for this team is there's 80 of them we have 14,000 employees right this card like how do you do this like really like its scale so this team in particular really looks at **how do they use the system you can call the bureaucracy to leverage and expand the value of the work** that they're trying to do right so it truly is one where it's like when this team kind of you're stood up in the parlance or lingo and the agency is um had a handful of folks coming to them with hey I got a problem can you help me solve it that kind of thing and looking at that and trying to be like okay well remember here **we're trying to solve problems but we're also trying to change the agency in the** culture right so like who are those like awesome rock star customers that you want to have in the consumer world I when I was studying this who was like **lead users even before early adopters**; **people who are doing crazy gonzo stuff** that you're like okay if we can meet there and eat and help them they are gonna evangelize for us they're gonna tell their friends do all those sorts of things so that's a lot of like how this team works like in the system really because a lot of people know they're very successful this team so you have a lot of people coming to you with like hey we want your help and assistance you got to kind of think strategically over like okay we got to solve this problem we got to solve this problem because if we get this one we can string together value for the agency right and I've kind of been a little big there and what I mean by value and I'll go more into that but it's really one where it's like how do we leverage the institution and the platform that we have to then scale the impact of what we're trying to create right you do it in this space and be very deliberate and be mindful of this we're trying to create and build the future of things for an agency like ours this is almost like the most critical part it's like **how do you understand the system to scale what you're trying to do who are those people that you need to get to those teams you need to** get to we talked earlier in our small group do you have the right color of money and that probably does it make any sense like if you're not from the DoD or the IC space but it's like money and funding and things like that are incredibly critical to figure out how you're going to scale impact right coming from the commercial world I was it was like nuts that you had different colors of money it wasn't being puppet anyway it's one of being aware of the systems and how they work trying to create this future can't just be about the experience of the data it is these systems as well and

1:43

here's where this team is really getting interesting to is much more in our within the work of this team because they're moving from that space of poll and stuff as spreadsheets and doing things so **looking much more like the implications of the work** that they're doing right and I can go into detail on that in the sense that it's like the implications of speeding up a workflow and making it that much faster because there is that thing where it's like well it used to take ten people I'm just making I'll just make up some numbers it's not a real **case it would take ten people two weeks to do one thing that would spit something out and now it takes like one person clicking on a script and they can do it in like ten minutes well like there's a lot of implications for that** **right with regard to employment and workforce and staffing** and billets and all that kind of stuff and billets just has to do with a number of people and the jobs and things like that again weird lingo but it's like looking at the implications of the work that they're trying to do so it's just one of in that sense of like how it impacts the workforce in our lingo we call it **tradecraft** really is like how you do your job right how does that impact those sort of things on one of the first order effects second order effects third or effects for this well it's been nice to with this team is they've been focusing on the implications at that level and now they've been able to get they've been quite successful in the work that they do **they're starting to look at and discuss implications or folks are looking to them for like what are the implications of doing data science and things like that in the entire intelligence community** so stakeholders are team members on this group our stakeholders on the **data science tradecraft initiative** other things that you might hear about like out here on the West or in the media about like hey what is the Department of Defense the intelligence community doing with data science and then they'll have like a working group and things like that these folks these team members are on these boards and panels guiding that what I love what that is because it's situated by folks who actually doesn't work right I mean that seems kind of ridiculous to say but it's one where it's like no it's like **you've got people who are in these conversations talking about the intersection of AI and ethics with save regard to counterterrorism we're about targeting and things like that where it comes from this like not just academic kind of discussion or model of like well what's the ethics of using automata to do some a certain type of work you've got folks who have done this right who have done it** for real whether themselves as well as situated with our workforce and our users to solve this sort of problem that what they're trying to do and

I think that's a fascinating work because this one where the implications part is one that really gets talked about like in the press **what are the implications of using artificial intelligence** for national security and things like that and this team is really driving and informing that dialogue and conversation with that real world here's what's work and here's what's not here's where we go

so I'll start wrapping here and it's just one it's like I think this framework is a friend we'll look at or it's a weight or a template to start looking at how do we build and create things right where do we sit ourselves in this kind of you can call it a ladder or a space you know to think about I'm trying to create X or Y am i living in the data space am i living in the technology space or where is that you know where am I and it's I think there's a tremendous lens for trying to figure out whether it's geospatial data science or data science in general it's like **how are we going to create what this future could be right how do we be very holistic in the work that we're trying to do how do we be very conscientious and deliberate about the actions that we take** because it's one of these that's like some of us would feel like we've been doing this kind of stuff for a long time it's still really early in the book right and like what we're trying to do is going to have tremendous implications or ramifications in the years going forward right I mean we forget that because we're too situated we're maybe in our like awesome innovation bubble of doing this cool work and things like that but **there's going to be a big ripple effect** so it's more of like how do we use these sorts of tools and frameworks to be very deliberate in the work that we're trying to do and going forward

1:47

so like **my action or my called really a for you like for OGC in the spac**e is one it's like going back to that diagram or you know riff your own make a new one where do you sit in that space like you look like where do you say right now like in your working in your job what would be the impact of going in a separate direction and going engaging with other sorts of folks how do you think holistically and the work that you're trying to do all right how do you go from the work that you are there to what is the minimal you know Minimal Viable Product did you do how can you start a conversation with other folks to start thinking about implications and things like that and then just jump in there and start doing it particularly from events like this which I think are tremendous work so you get folks together and you can nerd and rock out and you go like okay what is the one thing that we can do coming out of this right just to start moving us a little bit closer in this direction of creating what we want to be able to do with geospatial and data science in general so I'll wrap there just cuz I think we've got about 15 minutes hopefully we can get a little bit more on time happy to take any questions folks have about this prior work where the sources from this came from I'm gonna forgot to mention a couple credits on some of this didn't get on the slide in time but that's the Stanford slide was from Stanford d.school I got that from a tremendous event that I was at a couple weeks called the **machine learning and user experience meetup group** which is phenomenal like if you're in this space and you're here in the Bay Area it's a great group of folks who has a lot of practitioners who get together and just look at like and they really wrestle with great questions where it's like all right so we got some machine learning thing we're gonna put in some product what's the user experience research of that and like what we do about that I think they wrestle with those sorts of questions and it's lots of I would say like well-known but like just great people to interact with and this arose some of the slides and so the content arose a little bit from that there's a great community of folks to get together to help me try and figure out like what are these various things that we're working on in the agency and then provide this lens framework that it's ideally or hopefully useful to the public to understand what we're doing inside the inside the intelligence community which is very smoke and mioors often but also to as folks can take this and riff it and use it in different sorts of ways to go and create what they're trying to do with data science so I'll break there and then welcome any questions so thank you

**Questions**

1:49:26

Question: I found it interesting that the implications in your order came last can you explain a little bit more why experiences came first and implications came

Andy: sure are is that just literal flow of storytelling I think is one it's definitely one of those as you bounce around the way I describe it like with the team as it always starts with the experience you know for that team who is doing their work and I'll be clear two is data corps is one effort inside NGA who's doing this type of work we have other teams that I can give you word spaghetti over acronyms and things like that but there's other teams who are doing that I'm but you always **we always start with the experience part** is really that ground level of **what are you trying to do and how do you solve it** and sometimes you go to the implications very very quickly because you can quickly understand like oh this is their experience and if we do this one magical thing over here whoa like that can change a whole lot of stuff so **sometimes it will jump to that implications conversation**s before we even get to the data or the technology parts that was just you know challenges of linear storytelling

Question: I liked how you said that it was really valuable for you to have that kind of unform oh stop talk at the m/l event you went to and I wonder if that's a challenge within the intelligence community that you don't have those informal spaces to do that kind of technical kind of information sharing with other people in the geospatial community

Andy: I would say yes it is a challenge it is a barrier we have like informally **inside NGA our own kind of meetup groups** where folks will get together and like nerd out over coffee and things like that to do that that's a great first start it's one I'm really trying to push and we have advocates in the agency who are like I know folks at Google are wrestling with this right hard would you get into there's a cultural thing a lot of the folks that we have in our agency who have been there are like well I work in the IC I can't I don't want to list that I'm a public and this sort of thing so I don't want to engage or there's the perception to that like nobody wants to talk to the IC nobody wants to talk to us and those sorts of things and any time I hear those I was like you know just go find one person you'll find one person and you can engage in a conversation but I'm with you it's like **I would love to get more of the NGA folks and I'll even say CIA NSA and others who are wrestling with geospatial data and trying to figure out like what do we do engaging with I'll just call it folks outside of the IC** because most of really what we're running into and the challenges that we have have nothing to do with you being in the intelligence community nothing to do with national security or anything it's one of those where it's just like how do we get everybody to agree to a schema you know like which is legit you know like those sorts of things and training that sort of knowledge back and forth is key and critical we as an agency or we intelligent intelligence community can't rest on the oh we have to hide behind here because things are things are different or security and things like that know like the example I just gave is like now we should totally be engaging with others about how to do that so I'll often do is take it from that like oh we're scared of engaging with you know these folks out there is all brokering with colleagues from grad school and others and say like hey this group over here at Amazon and the MLS group is wrestling with that same question just give them all right and just start it like very low-level kind of conversations interactions **but I would love to have much I want to say formal but like a lot broader engagement with industry and others academia other folks who are wrestling with these sorts of challenges** so I think we'll all get smarter and better create that future by engaging with each other

Question: following up on that question how much of the data that you use is are just **open data** and I guess in terms of that you talked about experiences and developing solutions to them or you want walk through your process to do that are those again like internal process kinds of things or they also you're dealing with your broader issues of trying to apply many data sources to solving your broader yeah so so

Andy: the first question is um I can't answer that like specifically it's like what's the distribution you're like pie chart pie chart of you know you'll call it exquisite sensor data versus other sorts of things I can just say is like the data scientist is like I love more varied structured forms of data and there's tremendous sources available open that'll make our decision makers more informed so as one is I look at as us as an agency it's like more of that you know I would love to have more open data available to us why open is same reasons of open source or anything like that you got a lot of eyeballs well you know fill in all the blanks that kind of stuff would love to have more of that can you

Question: what was your second question one more time if you can okay or maybe was like the impact of the work that they're trying to do as a business process or external it's a mix of both

And: a lot of the work right now is the business process you know trying to clean up their tradecraft of how they do it not clean it up but it's make it more efficient and structured and documented

Question: I I like the way you sort of walked through sort of those particular steps and that resonated with being very prescriptive and so I guess I a question is do you ever sort of as an agency step through sort of that interaction with an agent with an agency that's a consumer really of your data and say we would normally tell you this is a better way to do this is more information you can gather this is a different methodology you ever step through and say I think your construct is wrong I think the future is this and not this so one example of that is changing the nature with nature of a way of the way an agent works though you know most folks as kind of said listen the agent is the gatekeeper we want to make them better what they do you ever challenged again

Andy: that's one where it's hard to answer that like super in specifics like with with because of classification stuff but I would say is yes like we challenging we push back I do yeah and the team does

Nadine question: you know the process that you described can you comment on whether you discovered or your team discovered **anything special about the geo part the location part** that makes it different from other domains not the Intel but like just actually the types of data that you're dealing with is there anything special there

Andy: I would say okay so one that I would step into is just like **how we store it has been a bit of a challenge** I don't mean just like putting in the cloud or something like that but it's just like how I might have to come back and answer that a second time like offline but it's like the first one that were just not offline but just clear the answer so to speak but one that I would think about is regard like what's makes it different with geospatial because I would say the data storage aspects of things and I can go into more detail I think following up with you of what that is but it's just not as straightforward as like when I worked in my physiological data and things like that yeah that's pretty straightforward we can store that this is different you know **whether it's the magnitude of the data or trying to tag it in different sorts of way it's just made a lot harder** what we're trying to do so it's a little bit of a duck but I guess got to be really careful okay thanks yeah

Question: so in yesterday's session there was a talk which i think was good in setting the landscape for the industry where there's a lot more data sets and types of the unit coming and someone mentioned metadata of data itself but there on the other hand on the processing side you've got hardware improvements with GPUs and then you've got software improvements with rewriting a Python script with C+ to have greater efficiencies **do you see if it's a race between there's always going to be more and more volumes of different types of data compared to the computing power** and processing and singing power that can do that or will the processing power eventually catch up and be able to compute it all this and the reason for that is from an analyst perspective we would shift or allocate our time from focusing on domain data and specialized datasets as opposed to focusing on improving the models and yeah

Andy: asked me to predict the future I feel like you have the best perspective on it's one word I look at it's like um **I get more and more excited by the data aspect of things** maybe that's cuz I'm less of an electrical engineer you know in like the computer hardware stuff is I always go it's like that'll be enough let's get smarter about what data is coming and how we use it and how its tags such that we can take advantage of the super fast GPU thing that's coming right that's that's where I focus more of my energy one is like I said the bias just from less familiarity with that but that's one where it's like this part is like my agency or my team's we have much more control but just kind of is like much more focus and directed to action on that we can impact that more than say waiting for the next supercomputer things like that because sometimes we'll get that it's like someone will come to us and say hey I got this data problem and my boss said that we should go use the supercomputer really like you know like what are you talking about yeah so it's like I always just try and table those like on the technology size just table that down a little bit and push it just because it's like that's just gonna keep going it's gonna keep growing faster let's focus our energy over here almost sometimes it's a little bit of all I'll say at a red herring right if we start focusing on those things and the power of compute that's coming is like we'll **wait what's the question you're trying to ask what is the decision what's the minimal data how is this going to make it because I know this is cool but it's kind of for us as an agency it's like kind of hard and expensive** not just us but is just like for the government's like hard and expensive to do all that kind of stuff so it takes a lot of work and I know it's got to be fun and cool but like what about over here you know what kind of things can we work on

Question online: question about open data and in particular follow-on to that comment on the impact of open data and the need for **higher quality data** do you struggle with acquiring acquiring higher quality data and what can be done to improve

Andy: yeah so that good question that's a really good one so struggling with the quality of the data that we've got coming in I know it's like struggling would be a hard hard I don't wanna use that language but its word it's one that we really look at it's like **for new data sources that are coming available regardless of where they are he's making sure that they have the provenance that they need that they have the structuring the labeling things** that they need we have a we call the commercial data office which I think looks at different data sources that we have which is public or paired with our we call it I think our **geoint assurance office** which kind of is like a formal infection or just instantiation of what I just described is like what's the provenance since data where did it come from and things like that those are very key strong efforts of what we have because one of the things that we really look at is **for any new data source that we have coming in how does it compare relative to what we've got right now** how will it potentially impact the decision making in tradecraft of folks that we have right now how do we catalog and log that and how do we illuminate that or share that with decision-makers who are you know making some high-stakes decisions just because of some great data well or some new data that's from available from open what really do we know about that right and how do we trace that and make that publicly available publicly in the sense to our analysts to our work force to guide their decision-making and then what I always love to do is get the feedback from them like how does this you know sniff test like how does this compare with what you think it should be right and **building those relationships and feedback loops with ideally the source of the data themselves** so even going back to commercial providers as much as we can you know and other sources and say hey well you know we saw it this way or we saw that is building that loop in

Question: you can for the presentation and you really enjoyed it we had a conversation yesterday on skill sets and career paths for geospatial and data scientists and I thought it was interesting that you said that the majority of your team is from outside of the Intel community yeah and I believe you said that your team goes out in teams of four sits with analysts and sees how data can help how they can help do you think **in that process the data corps people became more like analysts or the analysts became more like data scientists**

Andy: that's a really good question I don't know um I would say again anecdotal data you know kind of thing one of them all's we have like with this team is by bringing folks in from other places because we literally had people who are like hi gun amazing data scientists in the ad industry in Manhattan who are like wait what are you guys doing like that sounds awesome like that's an adventure I'll come join you and do those sorts of things who are now doing like amazing work I wish I could talk about but it's just like super cool focused stuff and you kind of goes like who do they learn and where do they **go well the model that we have with the Data corps and rolling it forward we bring those folks from outside in from industry get that paired process that experience with other folks and then ideally at a certain point they're like you know what I don't want to go back** I want to stay with this team like I jokingly or you know I use the reference of the team who's focused on what's going on with pickup truck all right okay so that'd be like the West Africa group or something and instead is taking those folks who have developed the skill and have this understanding and network with the analysts who are there and then become part of that team you know and really it's like do they become an analyst maybe they may become the data savvy analyst there who's training and educating and the way that we kind of look at is I use the **network contagion model but it's much more of like a hub-and-spoke kind of set up** as going is like cool you've been on this team for a while you've bounced around on this data corps team you've bounced around and solve problems in these different sorts of spaces what are you interested in and they're like you know what I'm really interested in building or working with this team over here you said great okay boom plug them over there and put them on that team to become more of an analyst right or just build out their analyst skillset so take all that like I said advertising and data background with the analyst data background some interesting stuff you can do put them together over here on that team and spread and grow so looking at very much like thinking of the team is like an organic like biological system you want the flow of people coming back and forth so you have the people going out but then also to is we have examples if we have analysts who are ones who are like brave enough to raise their hand and say like what is this team how can they help me in things like that and then all sudden they realize they're like they love metadata that's what they love to do you know they were an analyst but they're always kind of nerding out and they didn't feel like they were in their club and no one loved them or anything like that and you're like great so like come over here and be well we have data stewards which like if you love metadata this is the ultimate job for you know it's just like come and do that so that's happening as well which i think is great because what it is then is like there's someone who has been an analyst for like four years or seven years or whatever it is and just gets that like this role like how hard it is and what that job is like and now they're on this data team and they're the one who's like the expert or the more informed person than others so that's **really the model is getting that flow of folks in and out of the team itself both ways** yes I can't answer your question kind of answering a different question but hopefully I answered that net **we want that flow of folks to get both of those skillsets**

let's take a five or ten minute break and get the panel set up for that portion