ARL-TN-0931 ● Nov 2018

US Army Research Laboratory

Army Intelligence Data Science Workshop Summary Report

by Robert Ganger, John Coles, Jacob Ekstrum, Timothy Hanratty, Eric Heilman, Jason Boslaugh, and Zachary Kendrick

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TN-0931 ● Nov 2018

US Army Research Laboratory

Army Intelligence Data Science Workshop Summary Report

by Robert Ganger, John Coles, and Jacob Ekstrum
Information Exploitation, CUBRC, Inc., Buffalo, NY

Timothy Hanratty and Eric Heilman
Computational and Information Sciences Directorate, ARL

Jason Boslaugh and Zachary Kendrick
US Army Intelligence Center of Excellence

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
November 2018	Technical Note	1 June 2018–15 October 2018

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Army Intelligence Data Science Workshop Summary Report	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Robert Ganger, John Coles, Jacob Ekstrum, Timothy Hanratty, Eric Heilman, Jason Boslaugh, and Zachary Kendrick	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
US Army Research Laboratory ATTN: RDRL-CII-T Aberdeen Proving Ground, MD 21005	ARL-TN-0931

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Army Intelligence operates in a data-rich environment with limited ability to operationalize the exponentially increasing volume of disparate data. One concept being proffered as a solution to this challenge is the employment of data science. Toward this end, the US Army has funded an AR 5-5 study, titled "Application of Data Science within the Army Intelligence Warfighting Function (IWfF)", to determine how to employ future data science and data scientists to maximize data exploitation and reduce the burden on Army Intelligence Analysts. As part of this study, US Army Research Laboratory scientists, in collaboration with industry partner CUBRC and the US Army Intelligence Center of Excellence, led a two-day Data Science Workshop at George Mason University. The focus of the workshop was to gather experts from government, industry, and academia to discuss best practices and perspectives on utilizing data science within the IWfF. In addition to government representatives from US Army Intelligence and Security Command, US Army Training and Doctrine Command, US Army Forces Command, and US Army Special Operations Command, the workshop captured the experience of scientists from Microsoft, BAE Systems, RAND, and Lockheed Martin. This report summarizes the results of the workshop.

15. SUBJECT TERMS

Data Science, Military Intelligence, Intelligence Warfighting Function, Big Data, Analytics

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Timothy Hanratty
Unclassified	Unclassified	Unclassified	UU	43	19b. TELEPHONE NUMBER (Include area code) (410) 278-3084

Contents

1. Introduction

Army Intelligence operates in a data-rich environment with limited ability to operationalize the exponentially increasing volume of disparate data. The volume, velocity, and variety of data challenges the Army intelligence systems and processes, degrading the ability of the Intelligence Warfighting Function (IWfF) to provide timely, accurate, and relevant intelligence in support of mission command. One concept being proffered as a solution to this challenge is the employment of data science. Toward this end, the US Army funded an AR 5-5 study, titled "Application of Data Science within the Army Intelligence Warfighting Function", to determine how to employ future data science and data scientists to maximize data exploitation and reduce the burden on Army Intelligence Analysts. As part of this study, US Army Research Laboratory (ARL) scientists, in collaboration with industry-partner CUBRC and the US Army Intelligence Center of Excellence (USAICoE), led a two-day Data Science Workshop at George Mason University. The focus of the workshop was to gather experts from government, industry, and academia to discuss best practices and perspectives on utilizing data science within the IWfF. In addition to government representatives from the US Army Intelligence and Security Command (INSCOM), the US Army Training and Doctrine Command (TRADOC), the US Army Forces Command (FORSCOM), and the US Army Special Operations Command (USASOC), the workshop captured the experience of scientists from Microsoft, BAE Systems, RAND, Lockheed Martin, George Mason University, and University of Buffalo. This report summarizes the results of the workshop.

2. Workshop Format

Using a combination of large group discussions and breakout sessions, the workshop enabled all attendees to carefully consider the problem space and offer their expertise toward potential solutions. The collaborative and roundtable environment enabled the diverse group to form a consensus around a set of recommendations and priorities that will serve as critical information for the Army AR 5-5 study.

The participant breakdown was as follows:

- Active mMilitary: 10
- Government/contractor: 15

- Industry: 8
- Academia: 1
- Total participants: 34

To support the vision of the study to determine how to employ data science and scientists within the IWfF, the workshop was built around three breakout sessions. Each breakout session was led ("championed") by a subject matter expert (SME) and assigned a scribe to capture the session's findings. The champions for the sessions were CW5 Kevin Boughton, LTC Sam Huddleston, and Dr Aaron Frank.

2.1 Where Can Data Science Be Most Effectively Applied in a Military Intelligence Context?

Military Intelligence (MI) is performed at multiple levels (Brigade, Division, Corps, INSCOM), and covers many use cases including intelligence preparation for the battlespace, answering Priority Information Request/Commander's Critical Information Request (PIR/CCIR) questions, and battlespace situational awareness. Where in the chain of command and for what purposes is the creative work of data science best suited to enhance the intelligence process in the battlespace? In providing answers to this question, the team should consider issues such as the following and describe why their approach is valid:

- At what echelons should data science be employed? Are the data scientists in the field, or the continental United States (CONUS)? Is it a joint function?
- What intelligence activities are best served by data science? Requests for information (RFIs)? PIRs? CCIRs? Long-term studies? Targeting?
- What genres of analysis are best served? Imagery? Social networks? Social media? Studying adversary tactics, techniques, and procedures (TTPs)? Blue protection/safety? Event prediction?
- What (and why) are the short-, medium-, and long-term benefits of data science, compared against the cost of employing it? What can we do today for immediate results versus the desired future state (deeper analysis)?
- Will MI accept and respect data science as a standard intelligence practice? Will the results be trusted? If not in the short term, can results be introduced gradually to gain this trust?

2.2 What Are the Data Science Tools and Technologies That Will Benefit Military Intelligence?

The Army is investing in training data scientists to operationalize increasing volumes of both structured and unstructured data and to focus combat information so that the right unit receives the right information at the right time. However, the level of effort and potential payoff for using elements of data science in differing MI operational scenarios is highly variable, so we must identify and prioritize deployment of the *right* tools and techniques to bring measurable value. State of the art includes many tools and techniques for employing data science, including heuristics, machine learning, artificial intelligence, and big data analysis. The team will consider the known state of the art, including items in the following list, and recommend a short-, medium-, and long-term roadmap for gaining value quickly while planning for even greater gains in the future.

- What are the benefits versus costs/limitations of employing the following data science technologies?
    - Data warehousing/data engineering (conditioning, storage, accessibility, etc.)
    - Data mining/statistical analysis (exploratory work to reveal patterns, etc.)
    - Cloud and distributed computing (e.g., scalability, big data)
    - Database management/architecture (e.g., ontologies, schemas, data integration)
    - Business intelligence/strategy (e.g., dashboards, reporting, pedigree, documentation)
    - Machine learning/cognitive computing (e.g., automation, unsupervised learning, prediction)
    - Visualization (novel ways to present data and results)
- What are the benefits of employing data science tools versus the cost?
- What are the resource limitations for applying data science tools and techniques in MI?
- What tools and technologies get us to a measurable state of improvement quickly? Which ones are out of scope?

- What (and why) are the short-, medium-, and long-term benefits of applying these tools?

2.3 How Can Data Science Processes Be Leveraged to Benefit Military Intelligence?

The placement and employment of data scientists within existing doctrinal processes will be key to the success of the initiative. The data science field is inherently creative, which may not naturally mesh with the structured stepwise intelligence processes of the military. Thus, freedom for creativity must be balanced against priority deadlines and the battle rhythm. The team will assess data science practices and techniques against military intelligence processes to recommend how these two can be leveraged.

- Are there creative limits for data science processes in MI? What are they?
- How do we effectively fit data science processes into MI, both in terms of doctrine and field expediency?
- How do we encourage creative data science in the MI process-oriented setting?
- What are the major barriers to acceptance of data science processes? How could they be addressed?
- Is Agile a viable methodology for applying data science in a structured MI process? What would an Agile data science process look like?
- What (and why) are potential short-, medium-, and long-term approaches to incorporating data science processes into MI?

3. Findings

The following is a summary of the findings for each of the breakout sessions. The findings were agreed upon by each subgroup, presented to the large group by the session champion, and discussed in the full group to move toward consensus. Across the board, some of the key discussions and assessments of future data science work revolved around the following critical elements:

- **People First**: as education and personnel are the number one priority.
- **Timelines**: are often driven by the ability to acquire/educate/train talented people.

- **Technology**: is rapidly shifting and should not be the driving force behind long-term change.

Outlined in this section are the results presented by each group as part of their Session Outbrief. The outline is framed on priorities, timelines, risk, and assumptions. The combined results of the groups will be fused and used to support a broader USAICoE Doctrine, Organization, Training, Materiel, Leadership, Personnel, Facilities, Policy (DOTMLPF-P) study. Copies of the original Session Outbrief presentations can be found in Appendixes A, B, and C, respectively. Appendix D shows a summary chart of the tool and technique that should be deployed at each echelon.

3.1 Where Can Data Science Be Most Effectively Applied in a Military Intelligence Context?

3.1.1 Priority 1: Filling Gaps with Data Science (DS) Professionals

3.1.1.1 Near-Term Tasks

- Civilian DS Expert deployed to Military Intelligence Board (MIB-T)/ Tactical Special Operations Command (TSOC)/1st Sergeant First Class (SFC). Initial Gap fill is Open Source Intelligence (OSINT) – 12 (5-TSOC,5-MIB-T, 1-SFC, 1-USASOC) Person Test.
- Formalize partnership of Civilian DS with the new Domain DS (e.g., Data Science Domain Experts) – 2/year.
- Develop Private/Public partnership framework for internships/coop to accelerate DS capability:
  - Industry training with military.
  - Military training with industry.

3.1.1.2 Mid-Term Tasks

- Military DS Expert deployed to Expeditionary Military Intelligence Brigade (EMIB)/Psychological Operations (PSYOPS)/Rangers/Civil Affairs/ Special Operations Forces (SOF) Aviation command. Augment with Civilian Systems Engineering and Technical Assistance (SETA) contractors as needed.

- Physical/Virtual downward reinforcement of MIB-T DS personnel to the EMIB.
- Building 1–2 year fellowship/sabbatical/Intergovernmental Personnel Act (IPA) placements into the military for commercial personnel.
- Request additional resourcing to increase DS Domain Expert pipeline to 6/year.

3.1.1.3 Long-Term Tasks

- Have model users trained down to the Brigade Combat Team (BCT)/Special Forces Group (SFG) level.
- Explore direct commissioning opportunities for DS professionals.

3.1.2 Priority 2: Creating an Environment for DS Success

3.1.2.1 Near-Term Tasks

- Create positions at EMIB for DS.
- Explore how to develop 35T/S6 who can support data access and connectivity for DS.
- Coordinate efforts around DS tool and skill development across services.
- Work with the Process team to develop a faster process than RFI to leverage DS capabilities.

3.1.2.2 Mid-Term Tasks

- Keeping DS-trained people in DS roles, and ensuring that there is no negative effect on the advancement of the individual's career with a talent management process.
- Create a definition for a DS/Problem-solving Team.
- Leverage coordinated DS efforts across the services and support agencies.

3.1.2.3 Long-Term Tasks

- Deployment of a DS Team Concept of Operations (CONOPS) at MIB-T and EMIB (and SOF).

3.1.3 Priority 3: Building Capacity with Transferable DS Skills

3.1.3.1 Near-Term Tasks

- Training about data at all levels in Initial Military Training (IMT) and Professional Military Education (PME).
- Assess gaps and leverage the FA-49 Operations Research and Systems Analysis (ORSA) Track and/or Joint Defense Agencies (JDAs) to build DS capacity.
- Define what open or low-cost DS courses and skills the Army will pay for, and formalize the mechanism for support for credentialing opportunities online (COOL).

3.1.3.2 Mid-Term Tasks

- Providing real-world problems, data, and tools to people being trained for practice. Leverage an assumed collection management process to support continuous training and knowledge base development.
- Explore and integrate training for DS development for credentialing online.

3.1.3.3 Long-Term Tasks

- Data engineering/wrangling deployed separate from the DS and Intelligence and Electronic Warfare (IEW) (waiting on industry) down to BCT echelons.

3.1.4 Key Assumptions for "Where"

3.1.4.1 How Were Priorities Assessed/Ranked?

- Positively impacts mission readiness and outcome.
- DS investments fill an existing need.
- Build capability to fill future needs.

3.1.4.2 How Is Potential Payoff Measured?

- Prepares for near-peer adversary while also supporting sustained operations (e.g., 98% of expected future operations).
- Aligns with existing cost-benefit calculations for investment in MI, focusing on OSINT and supporting the All Source Analyst (35F/350F/35D).

3.1.4.3 Experienced Deployment

- The individuals running a model developed by DS professionals need to have some intuition as to whether or not the output is at least loosely in the right output space.
- Black boxing models in analytic processes means that if there is an error, the only way it is caught is if the person using the model has a reasonable notion of what an expected output should look like.
- Potentially limits the deployment of DS models down to the BCT level in the G2.

3.1.4.4 DS Can Buy Time: Best Place to Buy Time Is at the MIB-T Level

3.1.4.5 Include New Doctrine for DS in Army Doctrine

- Expand beyond Counter Insurgency–Counter Terrorism (COIN-CT) to support decision making used in operational environments where there was more time to fight an asymmetric threat.
- Must be adaptable to near-peer adversary combat.
- DS needs to be called out as a skill set that is delivered as a capability.

3.1.4.6 Incremental Approach to Capacity Building Is Key to Adoption of DS

3.1.4.7 The Development of Containerized Approaches to DS Deployment

- In order to maintain the security and integrity of a model, we assume that there will be a containerized deployment strategy to support roll-out of new models and DS tools.

3.1.4.8 Loss of Cutting Edge

- As a participant pointed out, "We are no longer on the cutting edge of products and services, so we must create public-private partnerships to accelerate innovation being leveraged in the Army."

3.1.4.9 Increase in PAI Utility

- There is an increasing body of information that is generated and accessible as Publicly Available Information (PAI) that will be critical to incorporate into the MI process.

- Example reference is the geospatial intelligence (GEOINT) sources that are/have transitioned from military sources (National Geospatial-Intelligence Agency) to PAI as a significant percentage of source material for analysis.

3.1.4.10 OSINT Is Critical for Future Success

- OSINT is currently identified as the best investment for MI, but there is no Military Occupational Specialty.
- Building the OSINT sources is a key need that DS could help support, as it is a component of each MI job.
- OSINT will allow the individual to build and maintain the right skill set.

3.1.4.11 Critical for DS to Partner with the Functional Disciplines

3.1.4.12 Access to DS Tools Will Be Provided at the Appropriate Levels

3.1.4.13 Access to Cloud-Architecture to Support Next-Generation DS Tools and Access

3.1.4.14 Collection Management Process Will Be Developed to Support Continuous Training and Knowledge Base Development

3.2 What Are the Data Science Tools/Techniques That Will Benefit the Military Intelligence?

3.2.1 Priority 1: Tool Push-Down

- Generate stable, immutable applications at higher echelons and allow lower level echelons to use these tools when relevant to the problem.
- "80% Now Solutions."

3.2.2 Priority 2: Cloud "Stack" Services

- Enable higher-level echelons to rapidly cycle through development and testing phases for new software (develop, test, production environments).
- Test-level technology should be available to lower echelons through cloud infrastructure (reach back).

3.2.3 Priority 3: Stack Consistency

- Develop a consistent, default set of programming languages to be used, such as R and Python, as well as consistency across the development environments and databases.
- Not all personnel must know all languages, but each relevant echelon should have enough diverse candidates to span all languages.

3.2.4 Payoff and Risks/Liabilities

3.2.4.1 Tool Push-Down

- Payoff:
  - Lower echelons can leverage data science, but are limited in their modification ability. This ensures that the lower echelon is using valid approaches developed at a higher echelon.
  - Fast prototyping: new tools can be quickly developed that can be discarded with minimal loss when new technologies become available.
- Risks/Liabilities:
  - Limited flexibility could prevent modification to tools by lower echelons when needed.

3.2.4.2 Cloud Stack Services

- Payoff:
  - Software can receive feedback and improve more rapidly.
  - New tools are more readily accessible to lower echelons during the earlier phases of each tool.
- Risks/Liabilities:
  - Using open-source tools as a base has security risks unless validated appropriately.

3.2.4.3 Stack Consistency

- Payoff:

- A standardized platform and environment will increase interoperability between all echelons (cloud architecture, division stacks, and local desktop/laptop).
- By having a diverse knowledge base at each echelon, each echelon can be expected to understand almost all tools with combined manpower.

- Risks/Liabilities:
  - More education (for data scientists) and training (for others) are required.
  - Limits the number of tools available if technologies become obsolete in the near term.

3.2.5 Key Assumptions for "What"

- Chart in Appendix A is viable for unit placement.
- Open-Source Data Stores/PAIs are available.
- Incoming Privates have a high school understanding of probability and statistics.
- Pull-down and Cloud stack solutions will involve reliable access to the cloud.
- Using an Agile-style process at the upper echelons.
- More people will be available at the division level in the future.

3.3 How Can Data Science Processes Be Leveraged to Benefit Military Intelligence?

3.3.1 Priority 1: Incorporate DS into Military Decision-Making Process (MDMP)

Reason: MDMP is the staff's decision-making framework, which the Intelligence Cycle (IC) supports and which can incorporate DS processes (e.g., Cross-Industry Process for Data Mining, Agile) to answer broad (e.g., sentiment) or focused questions (e.g., PIR).

3.3.1.1 Near-Term Tasks: Current MDMP/IC Does Not Need to Be Changed to Incorporate Data Science

- The process is sufficient.
- DS adds value by improving course of action development.
- Reduces analysis timeframe and/or improves confidence.
- DS team "Product Owner" is the Senior Intel Officer.
- Analysis is an input to decision making (not just an output of processes).

3.3.1.2 Mid-Term Tasks: Culture Change

- Break down data silos – to do multi-discipline analysis, need multi-discipline data.
- Data sharing is a necessity.
- Goal is to have a data lake – that enables data science to be performed.
- Change policies now to enable future changes.

3.3.1.3 Mid-Term/Long-Term Tasks: Push the Technology Further Down Echelon

- As you go down, the process is less human, and more technical.
- Tool and infrastructure evolution results in a process change (could do DS at lower levels with tools).
- Models produced higher up become "products" that are reused at lower levels.

3.3.2 Priority 2: Integrate Agile/Scrum/DevOps with Intelligence Cycle

Reasons: Improves communication; iteration improves product; Agile focuses on bringing product value early.

3.3.2.1 Near- to Long-Term Tasks

- Team design:
    - Product owner is the Senior Intel Officer.
    - Small interdisciplinary teams working together is best.
        - Analysts and computer scientists working as a team.

        - Leadership establishes open environment for team collaboration.
- Outstanding issues to resolve:
    - Pacing of analyses: DS team tempo needs to be in sync with requests.
    - Reachback: communication between distributed teams is a challenge; need trust up and down the chain.
    - Distinguish between one-off and ongoing analyses.
        - Enterprise level groups are needed to nurture the knowledge of ongoing analyses (e.g., OSINT sentiment of a country).
        - One-off analyses need to be prioritized and managed.
        - Process improvements, done and repeatable.
        - Contextually specific targeted analyses need to be scoped.
    - How to handle long-term tasking (e.g., a long-term analysis of a country does not fit in a sprint).
    - Tradeoff of high coordination/communication costs versus improved product from higher involvement of customers.

3.3.3 Priority 3: Data Governance

Reason: Data need to be discoverable and usable (data are an asset).

3.3.3.1 Short- to Long-Term Tasks

- Roles needed on the DS team: Intelligence Analyst, Data Scientist, Data Engineer, Data Journalist/Librarian, Data Communications/Visualizer.
- Need to be able to hand off data (i.e., match data to problems – across DS teams) and address questions around the management of that process.
    - Who is responsible for data translation? The DS team? The Intelligence Data owners?
    - Who hands data off in what form?
    - How to move and aggregate data?

- DevOps Environment: Need technology management model similar to Offensive Cyber Tools model – agile, quickly deployable, not bound by slow acquisition strategies.
- Quality Control:
    - Tool quality – OS tools with bugs; how to identify and fix issues.
    - Product quality – If a deployed product needs refinement, how to propagate updates and fixes.
    - Data quality/usage – Need inputs from data collectors to understand broader context of the data.

3.3.4 Priority 4: Importance of Investment in Human Capital

Reason: People are paramount, but developing, training, and retaining people is difficult. However, personnel expertise determines what analyses you can do at any level.

3.3.4.1 Near-Term Tasks

- DOTMLPF-P study will generate personnel requirements.
- Train our own as Military personnel – can get 10–15 years if you start early.
- Civilians – high turnover – is this a problem, or do we simply accept turnover as part of the culture?
- Contractors – [maybe] cost prohibitive.

3.3.4.2 Mid-Term/Long-Term Task

- Human capital market has highly uncertain future. Will market settle down?

3.3.5 Priority 5: Invest in Automation/DS in Targeted Areas

Reason: A phased-in approach for implementing DS at division or higher will enable value to be gained incrementally. For example, automated tools that simply visualize data in customizable ways may enable the human to spot the anomalies and issues, even if deeper DS analyses are not available.

3.3.5.1 Short- to Long-Term Actions

- Dedicated data science teams at Division or higher.

- Lowest level where it can be maintained as a dedicated resource.

- Data analyst – trained by DS in tools, could operate at lower level for smaller problems.
- A data scientist should be part of plan/direct to make sure the right questions are being asked.
- Areas where DS can help near-term:
  - Processing Exploitation and Dissemination
  - Alerting
  - OSINT
  - Data fusion
  - Tips and cues
  - Full Motion Video – reducing persistent stare

3.3.6 Key Assumptions for "How"

- DS in the near-term may generally be strategic rather than tactical.
- Resources get more constrained as you go lower in echelon, data channels are narrower, and analysis time is reduced.
- Commander is "customer" of DS product.
- Senior Intelligence Officer is "product owner" of DS product.

Appendix A. Session Group 1 WHERE Presentation

This appendix appears in its original form, without editorial change.

AMERICA'S ARMY:
THE STRENGTH OF THE NATION

UNCLASSIFIED

USAICoE Data Science Workshop

30-31 JUL 2018
Outbrief

Breakout #1:

Where *can data science be most effectively applied in a Military Intelligence context?*

Champion:

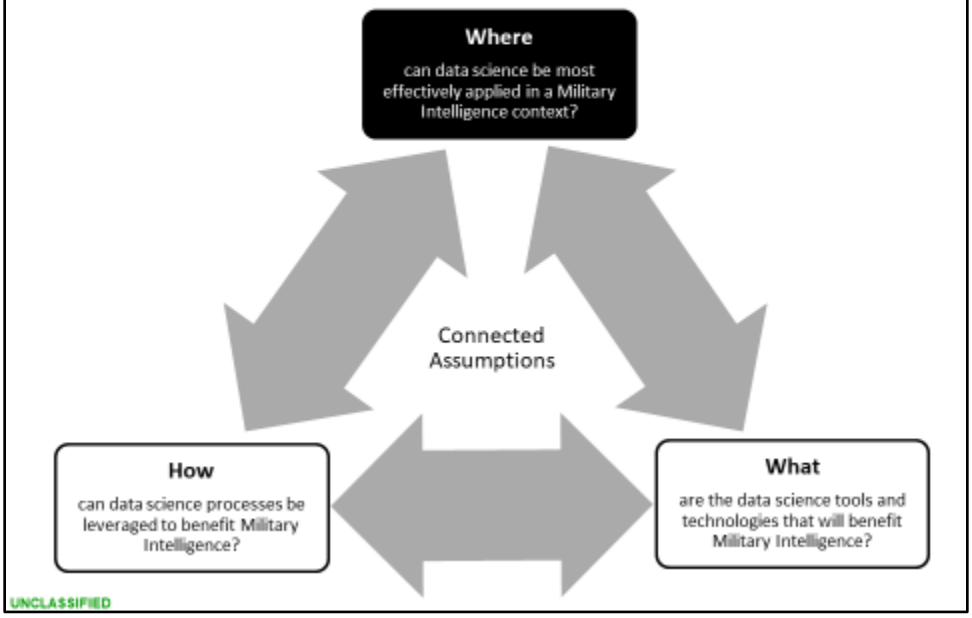
CW5 Kevin Boughton

UNCLASSIFIED

AMERICA'S ARMY:
THE STRENGTH OF THE NATION

UNCLASSIFIED

Breakout Context



UNCLASSIFIED

AMERICA'S ARMY: THE STRENGTH OF THE NATION

Mission for Breakout

Mission:

Identify the ***level of effort*** and ***potential payoff*** for using elements of data science in differing MI operational scenarios. ***Identify and prioritize*** deployment of data science in the right ***place*** using the right ***data*** at the right ***time*** to bring measurable value.

Outputs:

Generate a ***short, medium, and long-term roadmap*** for gaining value quickly, while planning for even greater gains in the future?

Outbrief:

Provide a summary of addressed **questions**, critical **assumptions**, **reference** material, and a **roadmap** for next steps.

UNCLASSIFIED

AMERICA'S ARMY: THE STRENGTH OF THE NATION

Priority 1: Filling Gaps with DS Professionals

1. Near-term Tasks:
    1. Civilian DS Expert deployed to MIB-T/TSOC/1st SFC. Initial Gap fill is OSINT – 12 (5-TSOC,5-MIB-T, 1-SFC, 1-USASOC) Person Test
    2. Formalize partnership of Civilian DS with the new Domain DS (e.g., Data Science Domain Experts) - 2/year
    3. Develop Private/Public partnership framework for internships/coop to accelerate DS capability
        1. Industry training with military
        2. Military training with industry
2. Mid-term Tasks:
    1. Military DS Expert deployed to EMIB/PSYOPS/Rangers/Civil Affairs/SOF-Aviation command. Augment with Civilian SITA contractors as needed.
    2. Physical/Virtual downward reinforcement of MIB-T DS personnel to the EMIB.
    3. Building 1-2 year fellowship/sabbatical/IPA placements into the military for commercial personnel.
    4. Request additional resourcing to increase DS Domain Expert pipeline to 6/year
3. Long-term Tasks:
    1. Have model users trained down to the BCT/SFG level.
    2. Explore direct commissioning opportunities for DS professionals.

UNCLASSIFIED

AMERICA'S ARMY: THE STRENGTH OF THE NATION

Priority 2: Creating an Environment for DS Success

UNCLASSIFIED

1. Near-term Tasks:
    1. Create Positions at EMIB for DS.
    2. Explore how to develop 35T/S6 who can support data access and connectivity for DS.
    3. Coordinate efforts around DS tool and skill development across services.
    4. Work with the Process team to develop a faster process than RFI to leverage DS capabilities.
2. Mid-term Tasks:
    1. Keeping DS Trained People in DS Roles, and ensuring that there is no negative effect on the advancement of the individual's careers with a talent management process.
    2. Create a Definition for a DS/Problem solving Team.
    3. Leverage coordinated DS efforts across the services and support agencies.
3. Long-term Tasks:
    1. Deployment of a DS Team CONOPS at MIB-T and EMIB (and SOF)

UNCLASSIFIED

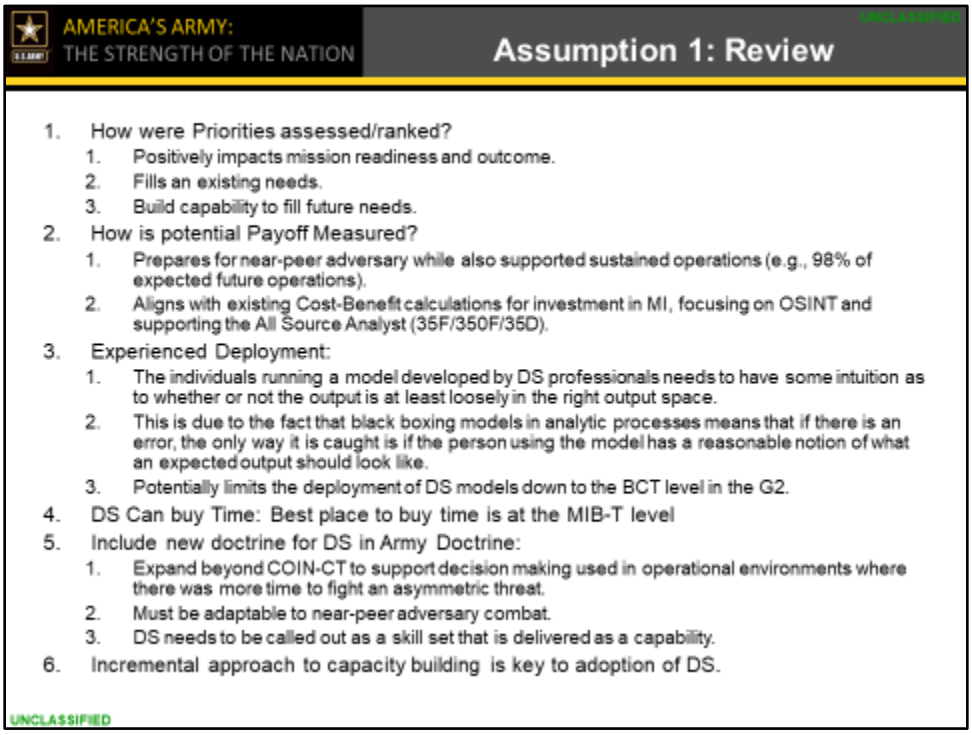
AMERICA'S ARMY: THE STRENGTH OF THE NATION

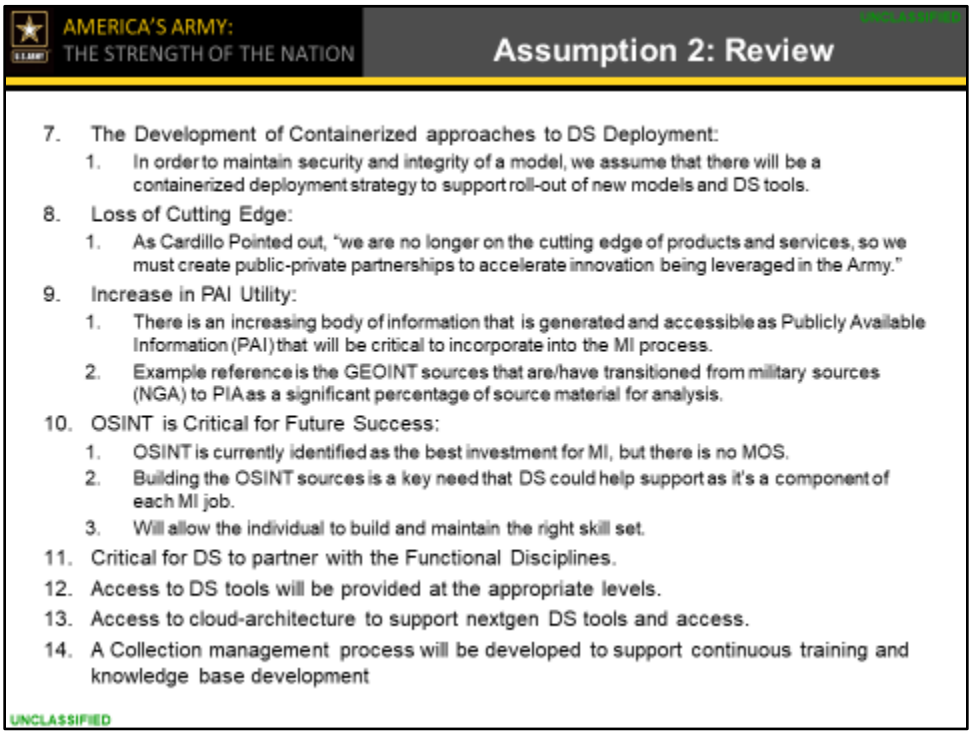
Priority 3: Building Capacity with Transferable DS Skills

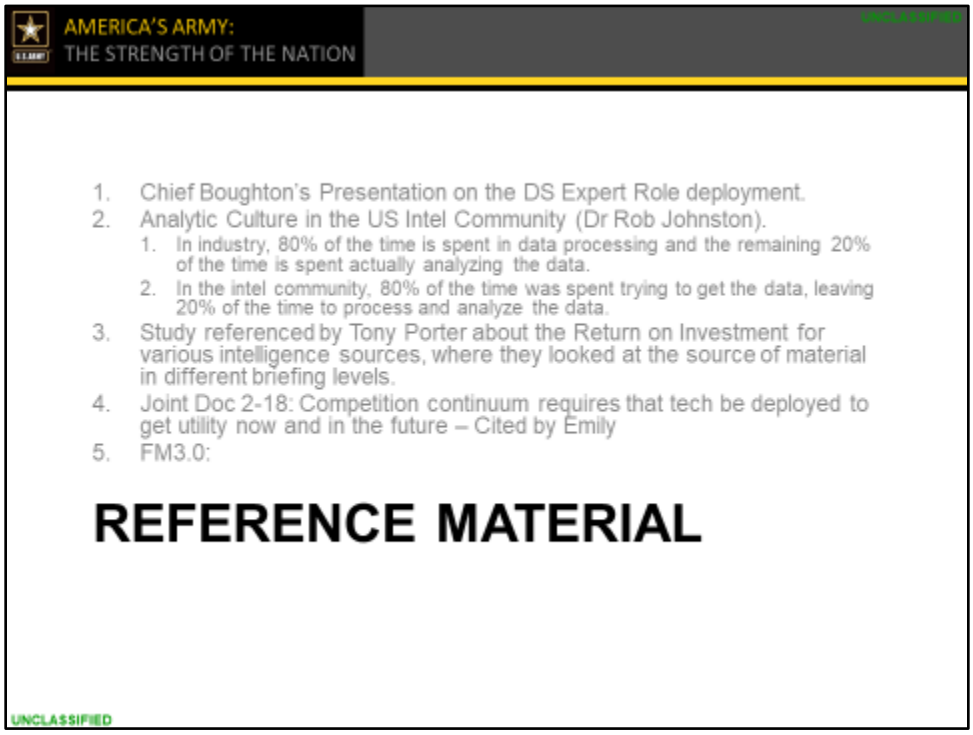
UNCLASSIFIED

1. Near-term Tasks:
    1. Training about data at all levels in IMT and PME.
    2. Assess gaps and leverage the FA-49 ORSA Track and/or JDAs to build DS capacity.
    3. Define what open or low-cost DS courses and skills the army will pay for, and formalize the mechanism for support for credentialing online (COOL).
2. Mid-term Tasks:
    1. Providing real-world problems, data, and tools to people being trained for practice. Leverage an ASSUMED collection management process to support continuous training and knowledge base development.
    2. Explore and integrate training for DS development for credentialing online.
3. Long-term Tasks:
    1. Data engineering/wrangling deployed separate from the DS and IEW (waiting on industry) down to BCT Echelons.

UNCLASSIFIED


AMERICA'S ARMY:
THE STRENGTH OF THE NATION
UNCLASSIFIED
Assumption 1: Review
1. How were Priorities assessed/ranked?
1. Positively impacts mission readiness and outcome.
2. Fills an existing needs.
3. Build capability to fill future needs.
2. How is potential Payoff Measured?
1. Prepares for near-peer adversary while also supported sustained operations (e.g., 98% of expected future operations).
2. Aligns with existing Cost-Benefit calculations for investment in MI, focusing on OSINT and supporting the All Source Analyst (35F/350F/35D).
3. Experienced Deployment:
1. The individuals running a model developed by DS professionals needs to have some intuition as to whether or not the output is at least loosely in the right output space.
2. This is due to the fact that black boxing models in analytic processes means that if there is an error, the only way it is caught is if the person using the model has a reasonable notion of what an expected output should look like.
3. Potentially limits the deployment of DS models down to the BCT level in the G2.
4. DS Can buy Time: Best place to buy time is at the MIB-T level
5. Include new doctrine for DS in Army Doctrine:
1. Expand beyond COIN-CT to support decision making used in operational environments where there was more time to fight an asymmetric threat.
2. Must be adaptable to near-peer adversary combat.
3. DS needs to be called out as a skill set that is delivered as a capability.
6. Incremental approach to capacity building is key to adoption of DS.
UNCLASSIFIED


AMERICA'S ARMY:
THE STRENGTH OF THE NATION
UNCLASSIFIED
Assumption 2: Review
7. The Development of Containerized approaches to DS Deployment:
1. In order to maintain security and integrity of a model, we assume that there will be a containerized deployment strategy to support roll-out of new models and DS tools.
8. Loss of Cutting Edge:
1. As Cardillo Pointed out, "we are no longer on the cutting edge of products and services, so we must create public-private partnerships to accelerate innovation being leveraged in the Army."
9. Increase in PAI Utility:
1. There is an increasing body of information that is generated and accessible as Publicly Available Information (PAI) that will be critical to incorporate into the MI process.
2. Example reference is the GEOINT sources that are/have transitioned from military sources (NGA) to PIA as a significant percentage of source material for analysis.
10. OSINT is Critical for Future Success:
1. OSINT is currently identified as the best investment for MI, but there is no MOS.
2. Building the OSINT sources is a key need that DS could help support as it's a component of each MI job.
3. Will allow the individual to build and maintain the right skill set.
11. Critical for DS to partner with the Functional Disciplines.
12. Access to DS tools will be provided at the appropriate levels.
13. Access to cloud-architecture to support nextgen DS tools and access.
14. A Collection management process will be developed to support continuous training and knowledge base development
UNCLASSIFIED



Appendix B. Session Group 2 WHAT Presentation

This appendix appears in its original form, without editorial change.

AMERICA'S ARMY:
THE STRENGTH OF THE NATION

UNCLASSIFIED

USAICoE Data Science Workshop

30-31 JUL 2018
Outbrief

Breakout #2:

What *are the data science tools and technologies that will benefit Military Intelligence?*

Champion:

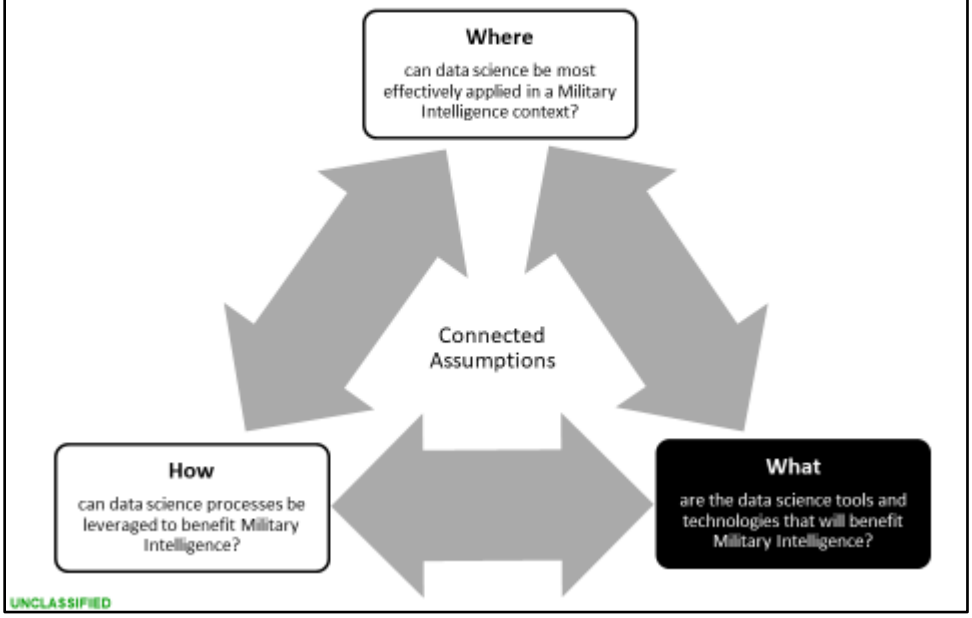
LTC Sam Huddleston

UNCLASSIFIED

AMERICA'S ARMY:
THE STRENGTH OF THE NATION

UNCLASSIFIED

Breakout Context



UNCLASSIFIED

AMERICA'S ARMY: THE STRENGTH OF THE NATION

UNCLASSIFIED

Mission for Breakout

Mission:

Identify the ***potential payoff*** for using elements of data science in differing MI operational scenarios. ***Identify and prioritize*** deployment of the right ***tools*** and ***techniques*** to bring measurable value.

Outputs:

Generate a ***near-term and future-oriented roadmap*** for gaining value quickly, while planning for even greater gains in the future.

Outbrief:

Provide a summary of addressed **questions**, critical **assumptions**, and a **roadmap** for next steps.

UNCLASSIFIED

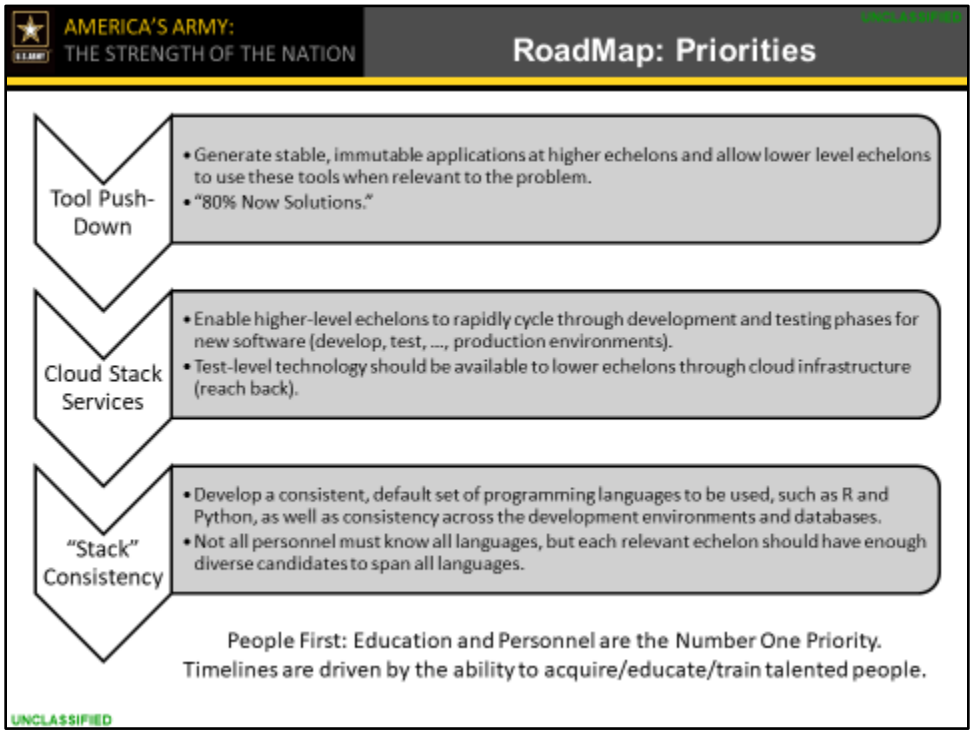
AMERICA'S ARMY: THE STRENGTH OF THE NATION

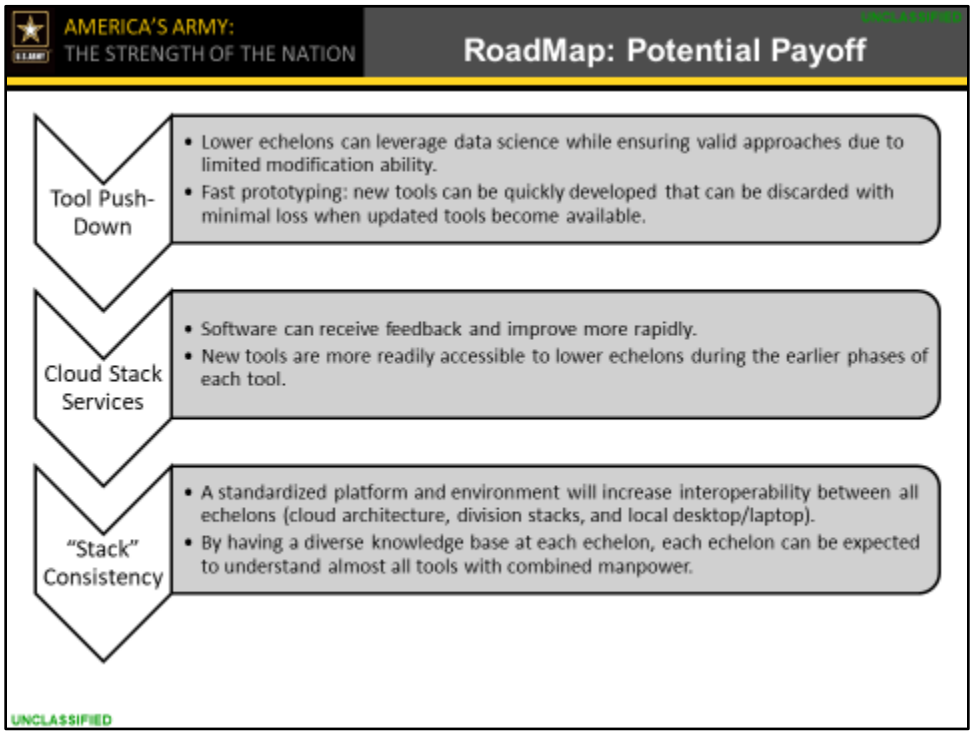
UNCLASSIFIED

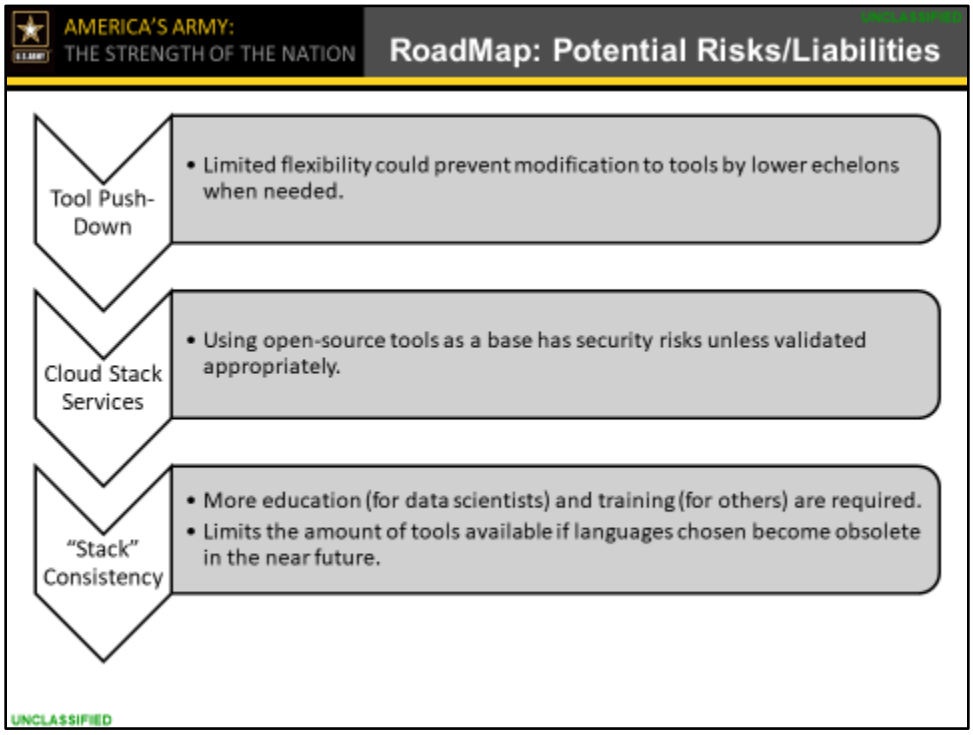
Discussion Summary

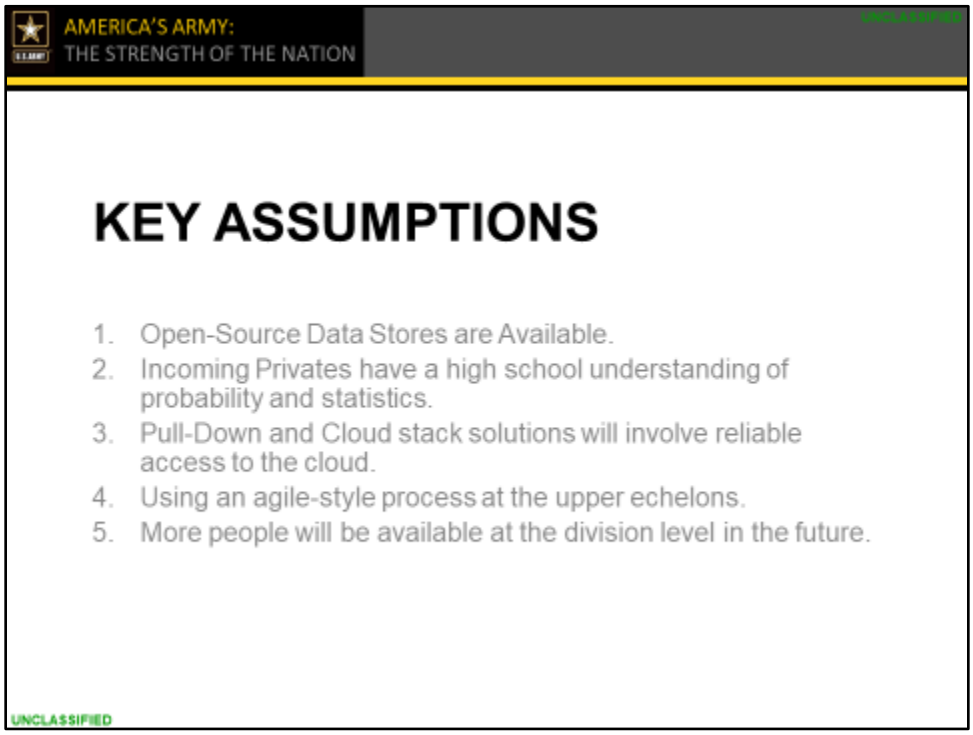
Time	Org/Level	Who is there?	What Data is there now?	What do they need?
>1 yr	INSCOM	• The goal – an MI Data Science Team	• Storehouses and Warehouses of regional data. • Can also pull from the agency level.	**Enterprise Data Science** • The ability to use a cyclical development cycle; develop and test in the cloud until ready for production. • Cloud stacks available that mimic those at the below echelons. • Cloud stacks available to test new technologies (databases, packages, modules, etc.). • The ability to host/deploy tools on servers and receive feedback. • Community boards. • On-call contracting support for scale, speed, and other demand-responsive resources. • Access to non-accredited tools. • The ability to develop and deploy containerized applications and environments.
72h-1yr	Corps	• O6 MI • MI E9 • CW4/5 • ORSA LTC	• Limited data warehouses. • Have access to GEOINT/SIGINT/FMV, etc.	**80 Percent Solution Now** • R/Python/Julia and other associated languages. • Streaming data support, such as Kafka. • MongoDB, SQL, and Graph databases. • Integrated Development Environments (IDEs). • Data analysis tools such as Tableau and Excel. • 3ST needs to know how to manage a stack and suite of programs (and on-call dashboards). • Open source tools, support for packages and modules.
24-72h	Division	• MI LTC • ORSA LTC • 2x ORSA MAJ • CW4 MI • MI E9	• Limited data warehouses. • Have DCGS-A database access.	
1-24h	Brigade	• Tool/App Users • End Users • MI MAJ • E8	• Any data that can be generated and stored on a laptop. • Can access higher data if needed.	**End User Applications** • Ability to use analytic applications on a laptop, either cloud-based or locally. • A compute platform that supports implementations of lightweight analytic tools, such as applications developed in R or Python. • Ability to receive and leverage containerized applications.
Now	Battalion	• EI 3SF • MI CPT • Tool/App Users • End Users		

UNCLASSIFIED


AMERICA'S ARMY:
THE STRENGTH OF THE NATION
RoadMap: Priorities
Tool Push-Down
• Generate stable, immutable applications at higher echelons and allow lower level echelons to use these tools when relevant to the problem.
• "80% Now Solutions."
Cloud Stack Services
• Enable higher-level echelons to rapidly cycle through development and testing phases for new software (develop, test, ..., production environments).
• Test-level technology should be available to lower echelons through cloud infrastructure (reach back).
"Stack" Consistency
• Develop a consistent, default set of programming languages to be used, such as R and Python, as well as consistency across the development environments and databases.
• Not all personnel must know all languages, but each relevant echelon should have enough diverse candidates to span all languages.
People First: Education and Personnel are the Number One Priority.
Timelines are driven by the ability to acquire/educate/train talented people.
UNCLASSIFIED


AMERICA'S ARMY:
THE STRENGTH OF THE NATION
RoadMap: Potential Payoff
Tool Push-Down
• Lower echelons can leverage data science while ensuring valid approaches due to limited modification ability.
• Fast prototyping: new tools can be quickly developed that can be discarded with minimal loss when updated tools become available.
Cloud Stack Services
• Software can receive feedback and improve more rapidly.
• New tools are more readily accessible to lower echelons during the earlier phases of each tool.
"Stack" Consistency
• A standardized platform and environment will increase interoperability between all echelons (cloud architecture, division stacks, and local desktop/laptop).
• By having a diverse knowledge base at each echelon, each echelon can be expected to understand almost all tools with combined manpower.
UNCLASSIFIED





Appendix C. Session Group 3 HOW Presentation

This appendix appears in its original form, without editorial change.

AMERICA'S ARMY:
THE STRENGTH OF THE NATION

UNCLASSIFIED

USAICoE Data Science Workshop

30-31 JUL 2018
Outbrief

Breakout #3:

How *can data science processes be leveraged to benefit Military Intelligence?*

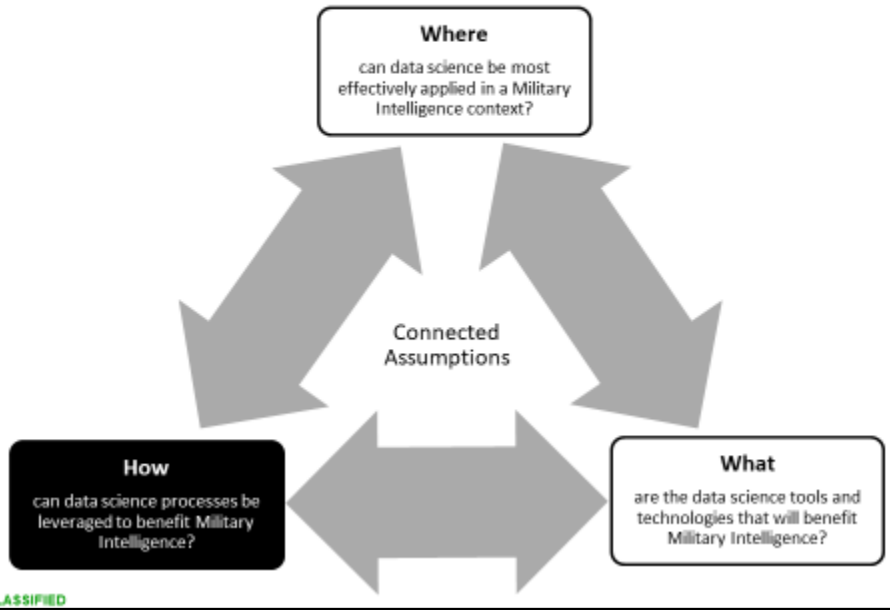
Champion:

Dr. Aaron Frank

UNCLASSIFIED

AMERICA'S ARMY:
THE STRENGTH OF THE NATION

Breakout Context



UNCLASSIFIED

AMERICA'S ARMY: THE STRENGTH OF THE NATION

Mission for Breakout

Mission:

Identify the ***level of effort*** and ***potential payoff*** for using elements of data science in differing MI operational scenarios. ***Identify and prioritize*** deployment of the right ***processes*** to bring measurable value.

Outputs:

Generate a ***short, medium, and long-term roadmap*** for gaining value quickly, while planning for even greater gains in the future?

Outbrief:

Provide a summary of addressed **questions**, critical **assumptions**, **reference** material, and a **roadmap** for next steps.

UNCLASSIFIED

AMERICA'S ARMY: THE STRENGTH OF THE NATION

Implementation Priority #1

- Military Decision Making Process (MDMP) →
  - Is the staff's decision-making framework which the Intelligence Cycle (IC) supports →
  - Which can incorporate data science (DS) process (e.g., CRISP-DM, agile) →
  - To answer broad (E.g., sentiment) or focused questions (E.g., PIR)
- Near-term priority: Current MDMP/IC does not need to be changed to incorporate data science
  - The process is sufficient
  - DS adds value by improving COA development
  - Reduces analysis timeframe and/or improves confidence
  - DS team "Product Owner" is the Senior Intel Officer
  - Analysis is an input to decision making (not just an output of processes)
- Mid-term priority: Culture change
  - Break down data silos – to do multi-discipline analysis, need multi-discipline data
  - Data sharing is a necessity
  - Goal is to have a data lake – that enables data science to be performed
  - Change policies now to enable future changes
- Mid/Long-term priority: Push the technology further down echelon
  - As you go down, the process is less human, and more technical
  - Tool and infrastructure evolution results in a process change (could do DS at lower levels with tools)
  - Models produced higher up become "products" that are reused at lower levels
- **Does Intel analysis and production merge with other disciplines and analysts (E.g., operational support, op-intel co-production)?

UNCLASSIFIED

AMERICA'S ARMY: THE STRENGTH OF THE NATION

Implementation Priority #2

- Integrate Agile/Scrum/DevOps with Intelligence Cycle
  - Improves communication
  - Iteration improves product
  - Agile focuses on bringing product value early
  - Product owner is the Senior Intel Officer
  - Small interdisciplinary teams working together is best
    - Analysts and computer scientists as a team
    - Leadership establishes open environment for team collaboration
  - Potential issues to solve:
    - Pacing of analyses: DS team tempo needs to be in sync with requests
    - Reachback: communication between distributed teams is a challenge; need trust up and down the chain
    - Distinguish between 1-off and ongoing analyses
      - Enterprise level groups needed to nurture the knowledge of ongoing analyses (E.g., OSINT sentiment of a country)
      - 1-off analyses
        - » Process improvements, done & repeatable
        - » Contextually-specific targeted analyses
    - How to handle long term tasking
      - E.g., a long term analysis of a country doesn't fit in a sprint
    - Tradeoff of high coordination/communication costs vs. improved product from higher involvement of customers

UNCLASSIFIED

AMERICA'S ARMY: THE STRENGTH OF THE NATION

Implementation Priority #3

- Data Governance
  - Roles needed on the DS team
    - Intelligence Analyst, Data Scientist, Data Engineer, Data Journalist/Librarian, Data Communications/Visualizer
  - Data needs to be discoverable and usable (data is an asset)
    - Data Journalist/Librarian needed
    - Need to be able to handoff data (I.e., match data to problems – across DS teams) – who is responsible for data translation? The DS team? The INT owners? Who hands it off in what form? How to move and aggregate it?
  - DevOps Environment
    - Need technology management model similar to Offensive Cyber Tools model – agile, quickly deployable, not bound by slow acquisition strategies
  - Quality Control
    - Tool quality – OS tools with bugs; how to identify and fix issues
    - Product quality – If a product goes out incorrect, how to propagate updates and fixes
    - Data quality/usage – Need inputs from data collectors to understand broader context of the data

UNCLASSIFIED

AMERICA'S ARMY: THE STRENGTH OF THE NATION

Implementation Priority #4

- Importance of Human Capital
  - People are paramount!
  - But, developing, training, retaining people is difficult
  - Personnel expertise determines what analyses you can do at any level
    - DOTmLPF-P study will generate personnel requirements
  - Near Term: Train our own
    - Military personnel – can get 10-15 yr if you start early
  - Civilians – high turnover – this is a problem
    - Or, do we simply accept turnover as part of the culture?
  - Contractors – [maybe] cost prohibitive
  - Mid-term/Long-term
    - Human capital market has highly uncertain future. Will market settle down?

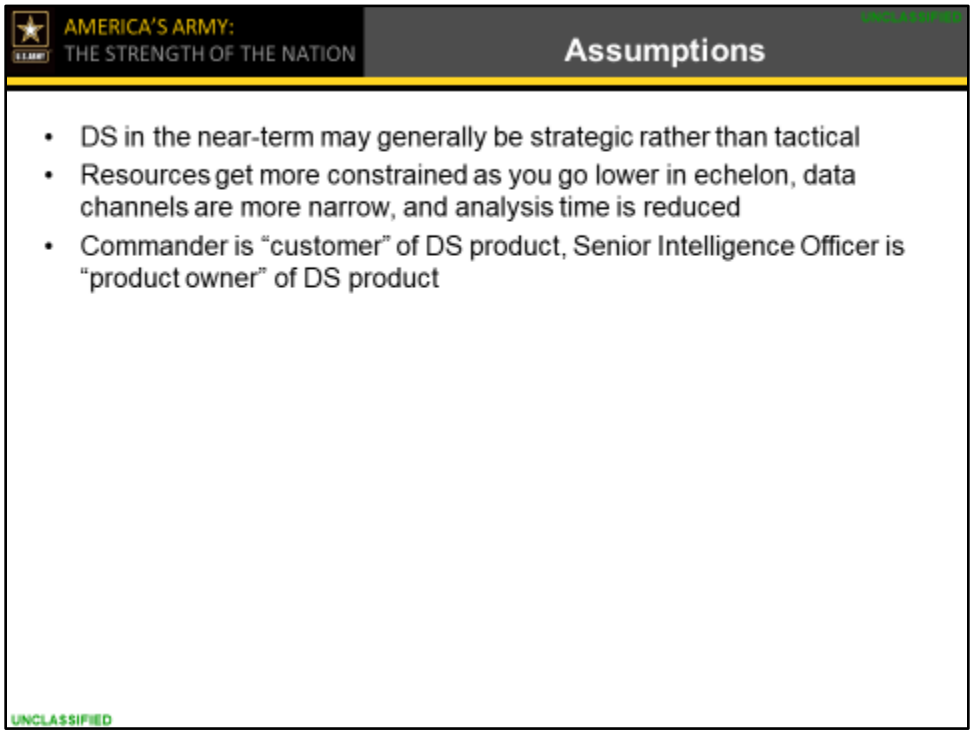
UNCLASSIFIED

AMERICA'S ARMY: THE STRENGTH OF THE NATION

Implementation Priority #5

- Best opportunities/places to invest in automation/DS
  - Need to visualize data so that the human can spot the anomalies and issues
  - Dedicated data science teams at Division or higher
    - Lowest level where it can be maintained as a dedicated resource
  - Should DS team become the data hoarder, and inform others of what data exists?
  - Data analyst – trained by DS in tools, could operate at lower level for smaller problems
  - A data scientist should be part of plan/direct to make sure the right questions are being asked
    - What is being asked?
    - Why is it being asked?
  - Areas where DS can help near-term
    - PED
    - Alerting
    - OSINT
    - Data fusion
    - Tips & cues
    - FMV – reducing persistent stare

UNCLASSIFIED



Appendix D. Summary Chart on Tools/Technique Needs at Different Echelons

This appendix appears in its original form, without editorial change.

Generated by the "What" breakout team in an effort to understand what tools and technologies should be deployed at each level.

Time	Org/Level	Who is there?	What Data is there now?	What do they need?
>1 yr	INSCOM	• The goal – an MI Data Science Team	• Storehouses and Warehouses of regional data. • Can also pull from the agency level.	**Enterprise Data Science** • The ability to use a cyclical development cycle; develop and test in the cloud until ready for production. • Cloud stacks available that mimic those at the below echelons. • Cloud stacks available to test new technologies (databases, packages, modules, etc.). • The ability to host/deploy tools on servers and receive feedback. • Community boards. • On-call contracting support for scale, speed, and other demand-responsive resources. • Access to non-accredited tools. • The ability to develop and deploy containerized applications and environments.
72h-1yr	Corps	• O6 MI • MI E9 • CW4/5 • ORSA LTC	• Limited data warehouses. • Have access to GEOINT/SIGINT/FMV, etc.	**80 Percent Solution Now** • R/Python/Julia and other associated languages. • Streaming data support, such as Kafka. • MongoDB, SQL, and Graph databases. • Integrated Development Environments (IDEs). • Data analysis tools such as Tableau and Excel. • 35T needs to know how to manage a stack and suite of programs (and on-call dashboards). • Open source tools, support for packages and modules.
24-72h	Division	• MI LTC • ORSA LTC • 2x ORSA MAJ • CW4 MI • MI E9	• Limited data warehouses. • Have DCGS-A database access.	
1-24h	Brigade	• Tool/App Users • End Users • MI MAJ • E8	• Any data that can be generated and stored on a laptop. • Can access higher data if needed.	**End User Applications** • Ability to use analytic applications on a laptop, either cloud-based or locally. • A compute platform that supports implementations of lightweight analytic tools, such as applications developed in R or Python. • Ability to receive and leverage containerized applications.
Now	Battalion	• EI 35F • MI CPT • Tool/App Users • End Users		

List of Symbols, Abbreviations, and Acronyms

ARL	US Army Research Laboratory
BCT	Brigade Combat Team
CCIR	Commander's Critical Information Request
COIN-CT	Counter Insurgency–Counter Terrorism
CONOPS	Concept of Operations
CONUS	Continental United States
COOL	Credentialing Opportunities Online
CTR	Contractor
DOTMLPF-P	Doctrine, Organization, Training, Materiel, Leadership, Personnel, Facilities, Policy
DS	Data Science
EMIB	Expeditionary Military Intelligence Brigade
FORSCOM	US Army Forces Command
GEOINT	geospatial intelligence
GMU	George Mason University
IC	Intelligence Cycle
IEW	Intelligence and Electronic Warfare
IMT	Initial Military Training
INSCOM	US Army Intelligence and Security Command
IPA	Intergovernmental Personnel Act IWfF Army Intelligence Warfighting Function
JDA	Joint Defense Agency
MDMP	Military Decision-Making Process
MI	Military Intelligence
MIB-T	Military Intelligence Board

ORSA	Operations Research and Systems Analysis
OSINT	Open Source Intelligence
PAI	Publicly Available Information
PIR	Priority Information Request
PME	Professional Military Education
PSYOPS	Psychological Operations
RFI	Request for Information
SFC	Sergeant First Class
SFG	Special Forces Group
SETA	Systems Engineering and Technical Assistance
SOC	Special Operations Commander
SOF	Special Operations Forces
TRADOC	US Army Training and Doctrine Command
TSOC	Tactical Special Operations Command
TTPs	Tactics, Techniques, and Procedures
USAICoE	US Army Intelligence Center of Excellence
USASOC	US Army Special Operations Command

1 (PDF) DEFENSE TECHNICAL INFORMATION CTR DTIC OCA

2 (PDF) DIR ARL
IMAL HRA
RECORDS MGMT
RDRL DCL
TECH LIB

1 (PDF) GOVT PRINTG OFC
A MALHOTRA

1 (PDF) ARL
RDRL CII T
T HANRATTY