



Leveraging U.S. Army Administrative Data for Individual and Team Performance

Presentation for:

**Administrative Data Research Facilities Network (ADRF)
Research Conference:
Innovations in Administrative Data**



November 14, 2018



Biocomplexity Institute & Initiative



Nathaniel J. Ratcliff, PhD
Kelly S. Ervin, PhD
*U.S. Army Research Institute
for the Behavioral and Social Sciences (ARI)*

Joshua Goldstein, PhD
Vicki Lancaster, PhD
Sallie Keller, PhD
Stephanie Shipp, PhD
*University of Virginia, Social and Decision
Analytics Division (SDAD)*

Joel Thurston, PhD
*Center for Army Leadership (CAL)
U.S. Army Combined Arms Center*



ARI Mission

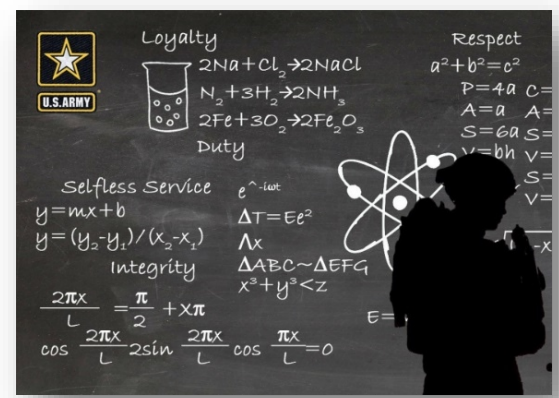
MISSION: Drive scientific innovation to enable the Army to acquire, develop, employ, and retain professional Soldiers and enhance personnel readiness.

Soldier-Oriented S&T



Develop innovative measures, methods, and models to maximize personnel and unit readiness of the Future Army.

Basic Research



Develop fundamental theories and investigate new domain areas in behavioral and social sciences with high potential impact on Army issues.

Scientific Assessments



Conduct scientific assessments and provide actionable findings to inform Army leaders and policy.



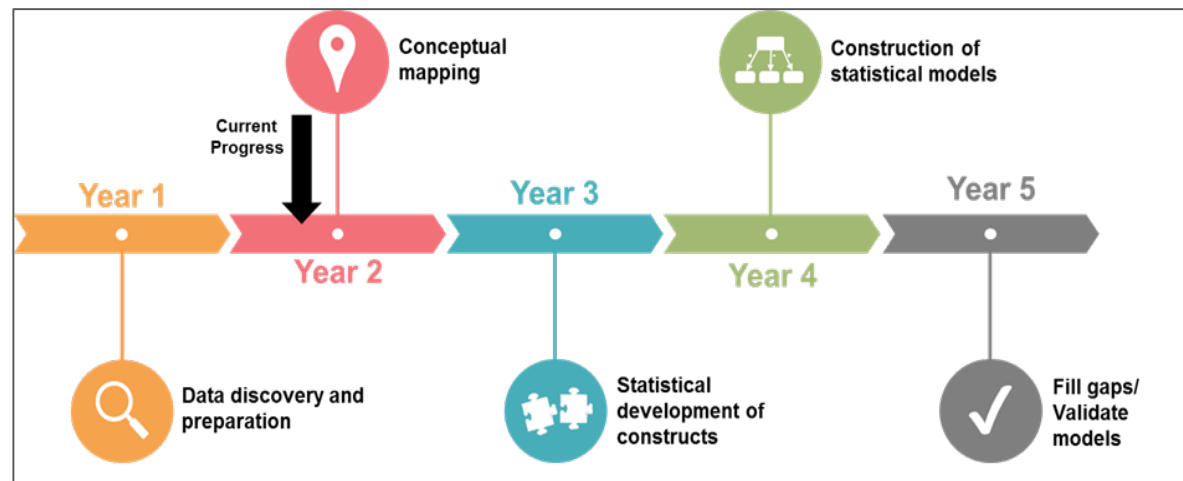
ARI: WHERE PERSONNEL SCIENCE MEETS PERSONNEL PRACTICE



The Problem Space

- **Problem:** The Army possesses a trove of administrative data (e.g., personnel records, training scores) but has yet to fully leverage these data
- **Purpose:** Use state-of-the-art analytics to develop models that integrate existing DOD data and make predictions about Soldier behavior and performance. This research will provide statistical models that forecast and visualize individual and unit performance
- **Payoff:** Knowledge about how best to utilize data from disparate sources to form a holistic picture of Soldier and unit performance

- Streamline/Reduce training
- Identify informative and efficient metrics of performance
- Optimize talent management decisions across Soldier lifecycle





Joshua Goldstein, PhD



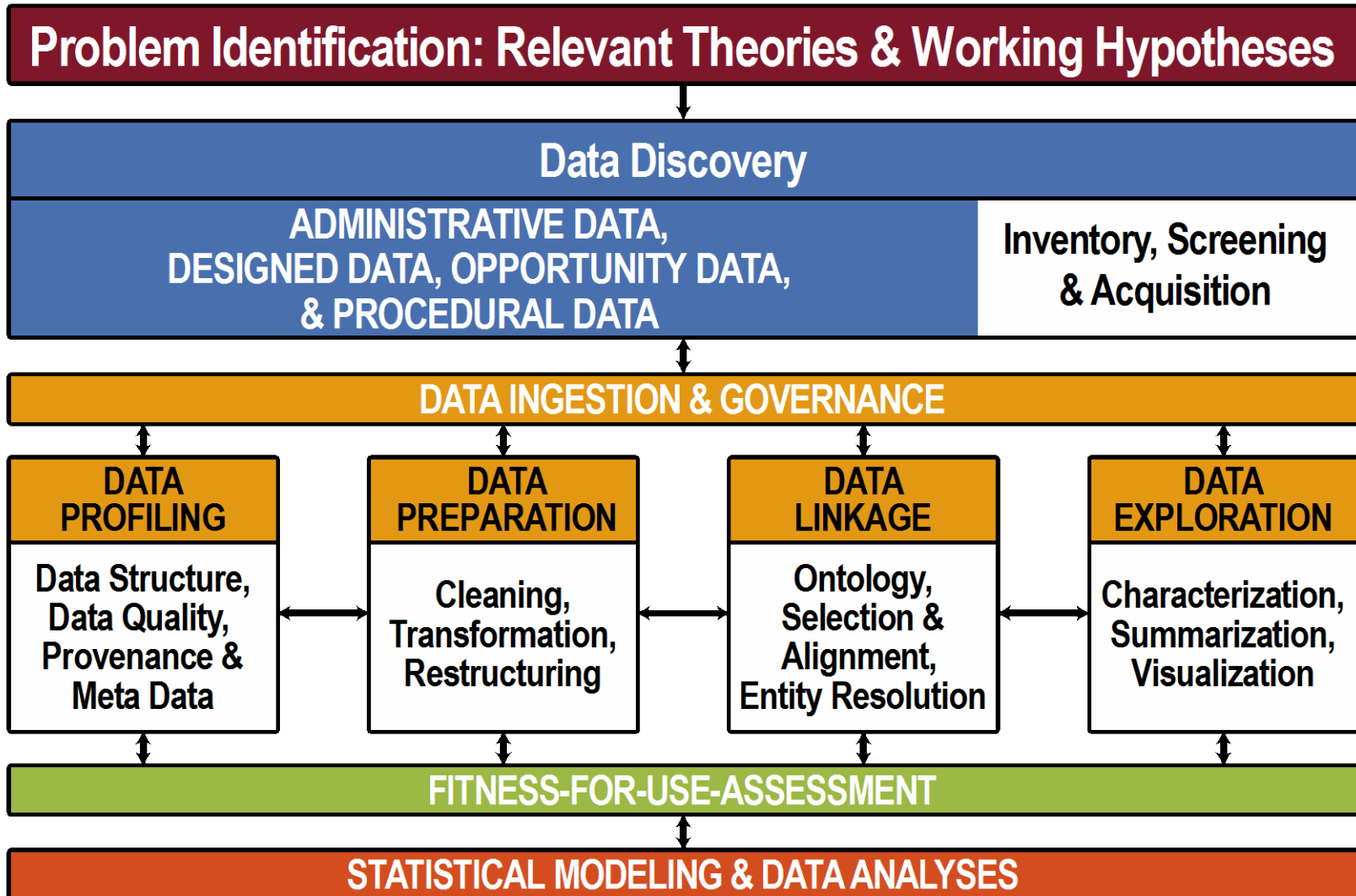
UNIVERSITY
of VIRGINIA

Biocomplexity Institute & Initiative





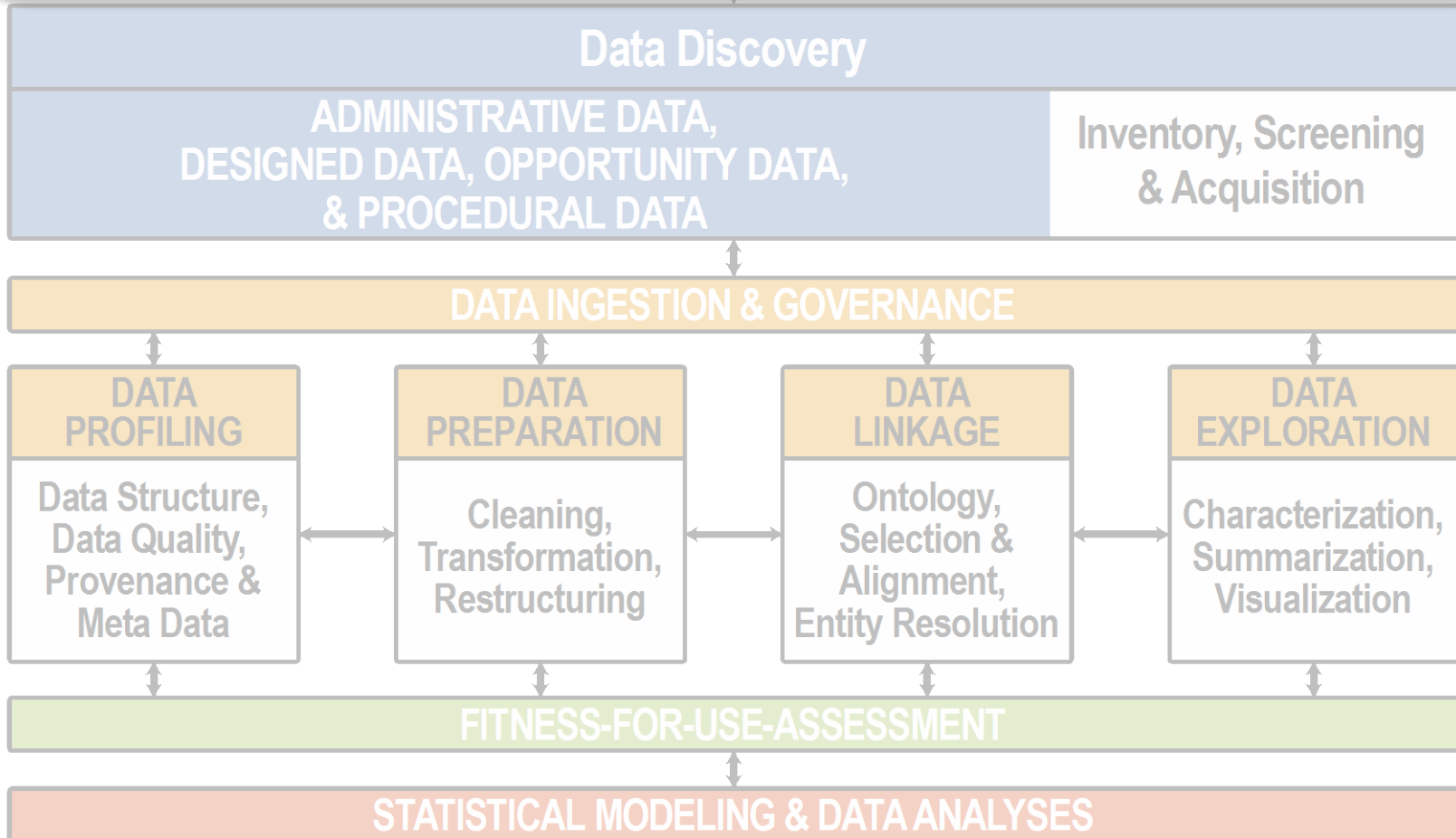
Data Science Framework





Data Science Framework

Problem Identification: Relevant Theories & Working Hypotheses





Data Science Framework

Problem Identification: Relevant Theories & Working Hypotheses

Data Discovery

ADMINISTRATIVE DATA,
DESIGNED DATA, OPPORTUNITY DATA,
& PROCEDURAL DATA

Inventory, Screening
& Acquisition

DATA INGESTION & GOVERNANCE

DATA
PROFILING

Data Structure,
Data Quality,
Provenance &
Meta Data

DATA
PREPARATION

Cleaning,
Transformation,
Restructuring

DATA
LINKAGE

Ontology,
Selection &
Alignment,
Entity Resolution

DATA
EXPLORATION

Characterization,
Summarization,
Visualization

FITNESS-FOR-USE-ASSESSMENT

STATISTICAL MODELING & DATA ANALYSES



Administrative Data Sources



Person-Event Data Environment (PDE)

- DOD maintains numerous datasets about military personnel including deployments, demographics, accessions/attrition, pay, promotions, awards, and training records
- Army Analytics Group Research Facilitation Lab (AAG-RFL) provides access to many of these datasets through the PDE in a secure, cloud-based enclave
- Accessing the PDE data housed at AAG-RFL requires:
 - A Common Access Card (CAC)
 - Human Research Protections/IRB approvals from all relevant institutions
 - Approval to access the PDE



Data Discovery – DOD Data Sources



Accessible in the PDE:

- Digital Training Management System
- Army Training and Requirements Resource System
- Individual Training History / DOD Formal Course File
- Army Workforce Transaction File: Military Awards
- Unit Risk Inventory
- Defense Equal Opportunity Management Institute (DEOMI) Organization Climate Survey
- Omaha 5 Behavioral Survey
- Active Duty Military Personnel Master
- Active Duty Military Personnel Transaction
- United States Military Entrance Processing Command (MEPCOM) Regular Army Analyst
- Interactive Personnel Elective Records Management System (IPERMS)
- Army Career and Alumni Program (ACAP)
- Pre-Deployment, Post-Deployment, Periodic Health Assessments
- Global Assessment Tool (GAT)



Data Discovery – Non-DOD Data Sources

Take a systematic approach in selecting from a huge inventory of public data sources and archives. Documented accessibility of data, geographies of interest, and relevant variables.

American Community Survey
Factfinder, Integrated
Public Use Microdata
Sample (IPUMS)

Aunt Bertha's

Bureau of Economic Analysis

Bureau of Labor Statistics

Quarterly Census of
Employment and Wages
(QCEW), Consumer Price
Index (CPI), Current
Employment Statistics
(CES), etc.

Community Commons

County Health Rankings

Gapminder World

Global Health Data Exchange

Headwater Economics

International Household Survey Network
Integrated Postsecondary Education
Data System (IPEDS) Postsecondary
Education

IRS Tax Records

Penn State Simple Online Data Archive

MIT Living Wage Calculator

National Center for Education Statistics

Population Reference Bureau

Simply Analytics

Statistical Atlas

U.S. Government Open Data

World Bank Open Data



Data Information Map

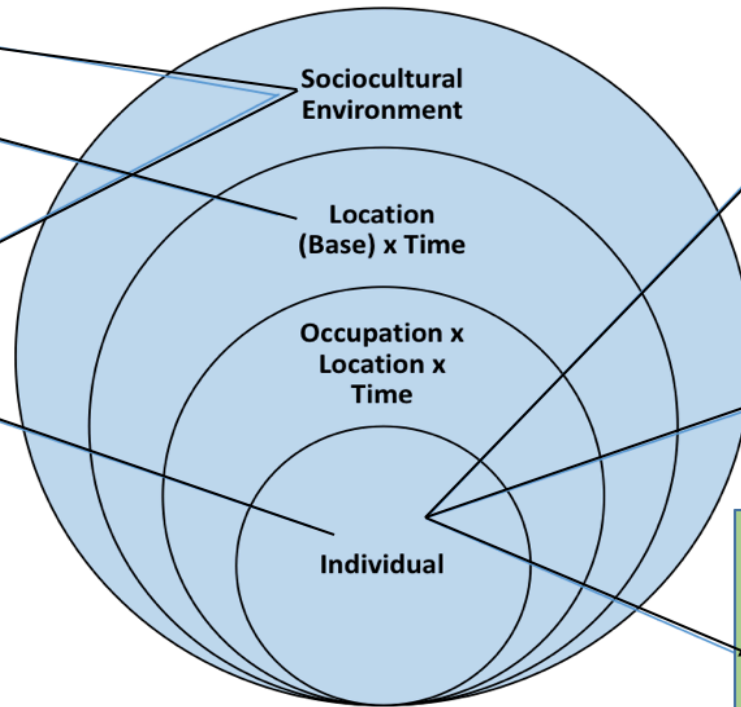
Data Needs to Operationalize the Antecedents and Social Determinants of Performance

Data Sources from the PDE with the potential to inform the performance model

Policy Changes (e.g., DADT)
 Non-personal shock events (e.g., 9-11)
 Job Opportunities
 Non-base Community

Constructs to Model

Job satisfaction
 Job Embeddedness
 Community Embeddedness
 Motivation
 DoD Ethical & Army Values
 Warrior Ethos
 Vocational Interests
 Creativity
 Organizational Citizenship
 Character Values
 Individual Differences
 Grit



Demographics

Race, Sex, Ethnicity, Birthdate, Age, Faith group,
 Education level and discipline, Marital status, Spouse in military indicator, Number and type of dependents, State/county of residence before entry, ASVAB score, APFT score, G.A.T. results

Service Dates and Locations

Length of time in service
 Length of service agreement
 Location (base) over time
 Obligation begin and end dates
 Term of service
 Date of initial entry
 Date of end of initial training

Military-Specific Characteristics/Incentives

Security Clearance
 Education Incentive Indicator
 Career status bonus program indicator
 Occupation group
 Re-enlistment eligibility
 Article 15
 Training schedule and training awards
 Rate of promotions
 Page grade / length of time in grade
 Character of service



Steps in the Data Quality Analysis Process

Completeness

Proportion of Elements Properly Populated for a Given Purpose

- Issue types include: Record fields containing no data; records not containing necessary fields; datasets not containing the requisite records (e.g., testing for NULLs and empty strings existing where not appropriate)

Validity

Proportion of Elements whose Attributes Possess Proper Values

- Checking for value validity generally comes in the form of straight-forward domain constraint rules (e.g. where gender is not one of (male, female), or where age is not between [0, 110])

Uniqueness

Count of Unique Values Taken by an Attribute or Combination of Attributes

- Frequency distribution of an element. (Note: The more homogeneous the data values of an element, the less useful the element is for analysis)

Duplication

Degree of Replication of Distinct Observations Per Observation Unit Type

- For example, greater than one registration per student per official reporting period.
- Note: Duplication occurs as a result of choice of level of aggregation

Consistency (Record)

Degree to which two or More Data Attributes Satisfy a Dependency Constraint

- Relationship validation
- For example, zip-code – state consistency; gender – pregnancy consistency

Consistency (Longitudinal)

Degree an Attribute, Combination of Attributes, Remains Consistent Over Time

- For example, an individual's gender changing and then changing back (clerical error) or an individual's race classification changing from one to two races (change in number of options)



Demographic Log Analysis & Reduction



Basic demographics between data files need to be reconciled and reusable data products created.

Demographics Table

- Information about the enlistee that typically remains static over time (e.g., gender, race, ethnicity, entry test scores)
- Simple rules are applied to resolve duplicates and invalid entries
- Contains one row per personal identifier (PID)

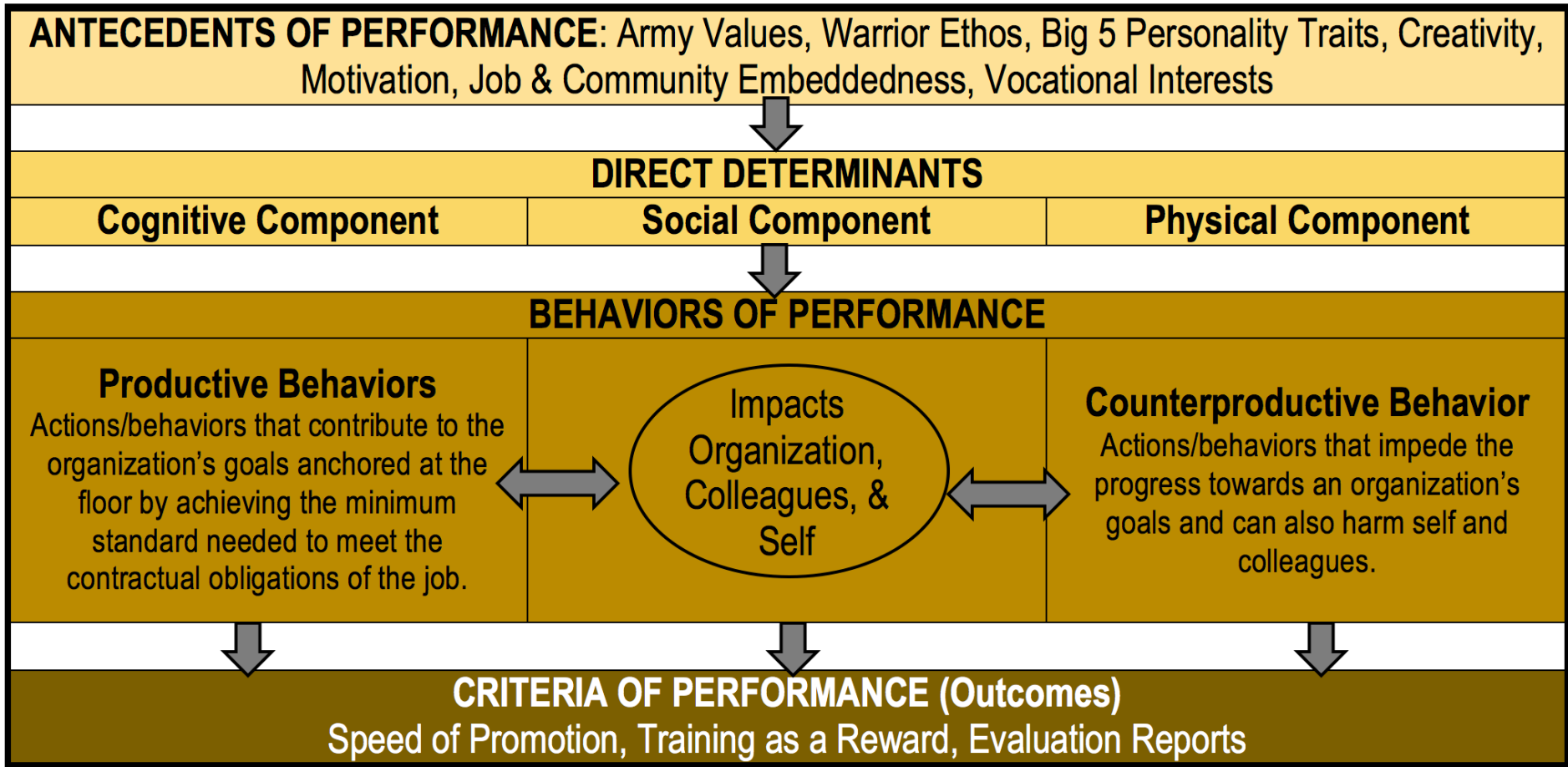
Transaction Table

- Events or enlistee information that can change periodically (e.g., duty station, rank, pay grade)
- Contains multiple rows per PID

Column Name	Description	Original Table
PID_PDE	Enlistee's Unique ID	Master
PN_SEX_CD	Gender	Master
RACE_CD	Race Code	Master
INIT_ENT_TRN_END_DT	Initial Entry Training End Date	Master
DATE_BIRTH_PDE	Person Birth Date	Master
PN_BIRTH_PLG_CTR_Y_CD	Person Birth Place Country Code	Master
HOR_ZIP_CODE_PDE	Home of Record Zip Code	Analyst
ACT_SCORE	ACT Score	Analyst
SAT_SCORE	SAT Score	Analyst
AP	ASVAB: Auditory Perception Score	Analyst
CO	ASVAB: Combat Score	Analyst
.	.	.
.	.	.
.	.	.

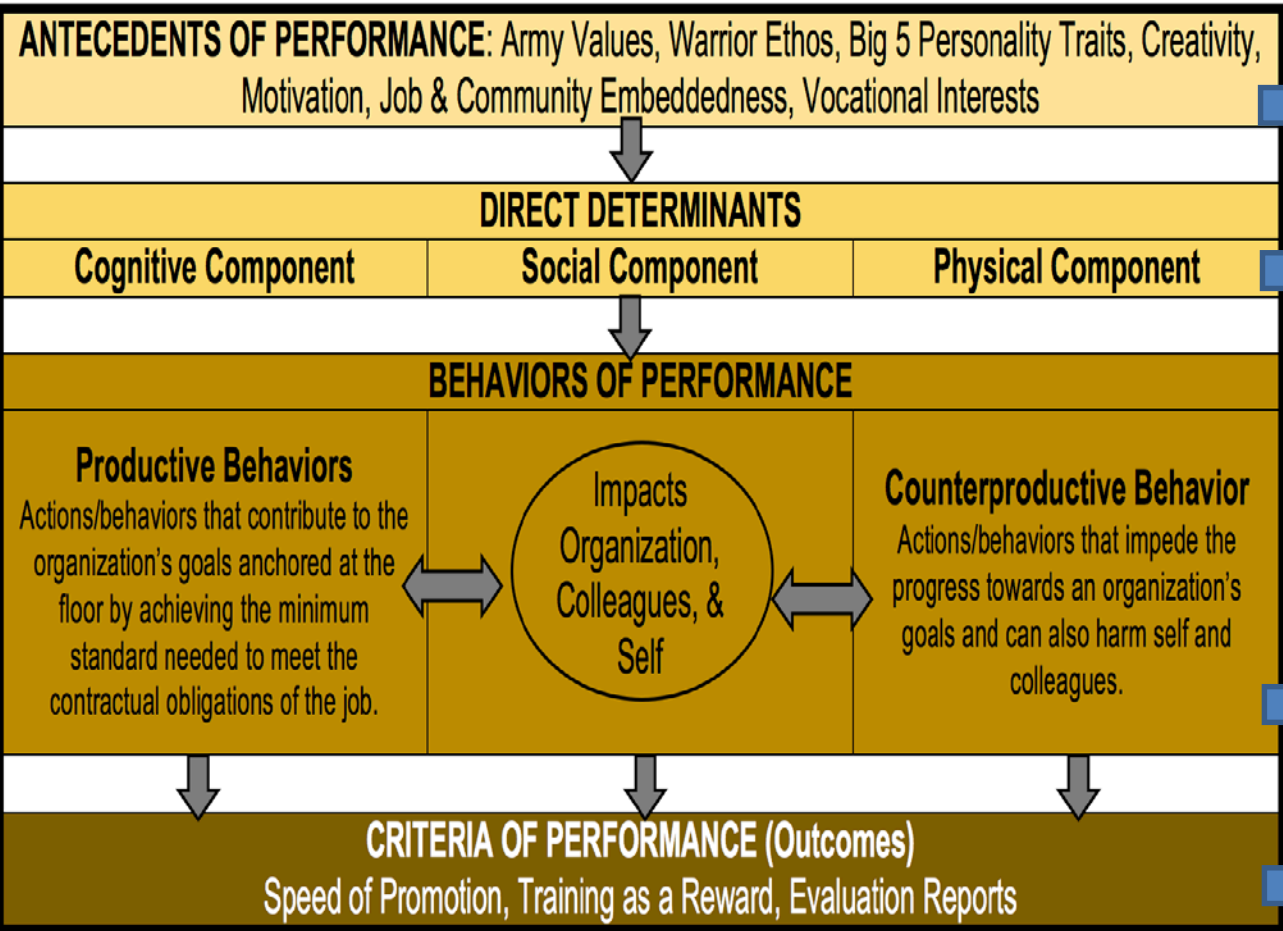


Conceptual Model of the Social Characteristics of Performance





Linking Data to the Performance Framework



Global Assessment Tool
External Community Data (e.g. Army Community Service (ACS), BLS

Armed Services Vocational Aptitude Battery (ASVAB) scores
Global Assessment Tool (GAT)
DEOMI Command Climate Survey
Unit Risk Inventory
Digital Training Management System (DTMS) Height and Weight
DTMS Army Physical Fitness Test

IPERMS Courts Martial
Article 15 Proceedings
Letters of Reprimand

Army Transaction File
Army Training and Resource Requirement System
Individual Training History



Questions

Contacts:

Kelly S. Ervin, PhD

U.S. Army Research Institute
for the Behavioral and Social Sciences
Kelly.s.ervin.civ@mail.mil

Joshua Goldstein, PhD

University of Virginia, Social and Decision
Analytics Division (SDAD)
joshua.goldstein@virginia.edu



We acknowledge the significant contributions of all co-authors on this project:

Nathaniel Ratcliff, PhD

U.S. Army Research Institute for the Behavioral and Social Sciences

Sallie Keller, PhD

Vicki Lancaster, PhD

Stephanie Shipp, PhD

University of Virginia, Social and Decision Analytics Division (SDAD)

Joel Thurston, PhD

Center for Army Leadership (CAL) U.S. Army Combined Arms Center

REPORT DOCUMENTATION PAGE			<i>Form Approved</i> OMB No. 0704-0188		
1. REPORT DATE (DD-MM-YYYY) 14-11-2018		2. REPORT TYPE Final		3. DATES COVERED (From - To) 30-6-2017 – 14-11-2018	
4. TITLE AND SUBTITLE Leveraging U.S. Army Administrative Data for Individual and Team Performance			5a. CONTRACT NUMBER W911NF-17-2-0116		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Joshua Goldstein (Social and Decision Analytics Laboratory), Vicki A. Lancaster (Social and Decision Analytics Laboratory), Nathaniel J. Ratcliff (U.S. Army Research Institute), Joel Thurston (U.S. Army Research Institute), Kelly S. Ervin (U.S. Army Research Institute), Daniel Chen (Social and Decision Analytics Laboratory), Ian Crandell (Social and Decision Analytics Laboratory), Bianica Pires (Social and Decision Analytics Laboratory), Aaron Schroeder (Social and Decision Analytics Laboratory), Stephanie Shipp (Social and Decision Analytics Laboratory), and Sallie Keller (Social and Decision Analytics Laboratory)			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Social and Decision Analytics Laboratory (SDAL), Biocomplexity Institute of Virginia Tech Virginia Polytechnic Institute and State University, Virginia Tech National Capital Region 900 North Glebe Road Arlington, VA, 22203			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Institute for the Behavioral & Social Sciences 6000 6 TH Street (Bldg. 1464 / Mail Stop 5610) Fort Belvoir, VA 22060-5610			10. SPONSOR/MONITOR'S ACRONYM(S) ARI		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT: Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES. Presentation for the Administrative Data Research Facilities (ADRF) 2 nd Annual Research Conference, Washington, D.C., November 14, 2018					
14. ABSTRACT The Army possesses vast amounts of administrative (archival) data about Soldiers. These data sources include screening tests, personnel action codes, training scores, global assessments, physical fitness scores, and more. However, the Army has yet to integrate these data to create a holistic operating picture. Our research focuses on repurposing Army administrative data to (1) operationalize social constructs of interest to the Army (e.g., Army Values, Warrior Ethos, Character) and (2) model the predictive relationship between these constructs and individual (i.e., Soldier) and team (i.e., unit) performance and readiness. The goal of the research is to provide people analytics models for the purposes of optimizing human capital management decisions. This presentation describes the theoretical underpinnings of our human performance model, drawing on disciplines such as social and industrial/organizational psychology, as well as our experience gaining access to, and working with Army administrative data sources. Access to the archival administrative data is provided through the Army Analytics Group (AAG), Person-event Data Environment (PDE). The PDE is a business intelligence platform that has two central functions: (1) to provide a secure repository for data sources on U.S. military personnel; and (2) to provide a secure collaborative work environment where researchers can access unclassified but sensitive military data.					
15. SUBJECT TERMS Data, data administrative, archival data, DOD data, repurposed data, Army values, Warrior Ethos, Army character, Predictive Analytics, performance, readiness, unit performance.					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 19	19a. NAME OF RESPONSIBLE PERSON Colanda Cato, PhD
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER 703-545-2354