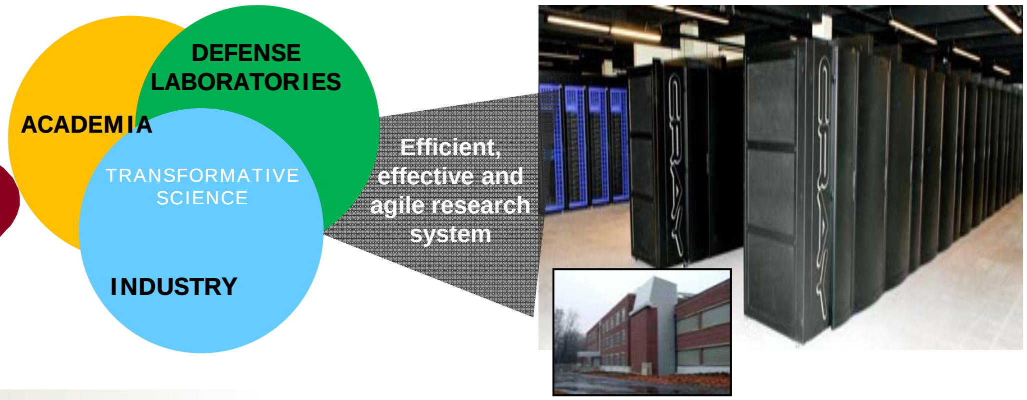


Data Intensive Science and Computing

University of Tennessee
&
Oak Ridge National Laboratory

October 29-30, 2015

TECHNOLOGY DRIVEN. WARFIGHTER FOCUSED.

Dr. Michael Barton
Computational Sciences Division
U.S. Army Research Laboratory

Our Mandate

Vision

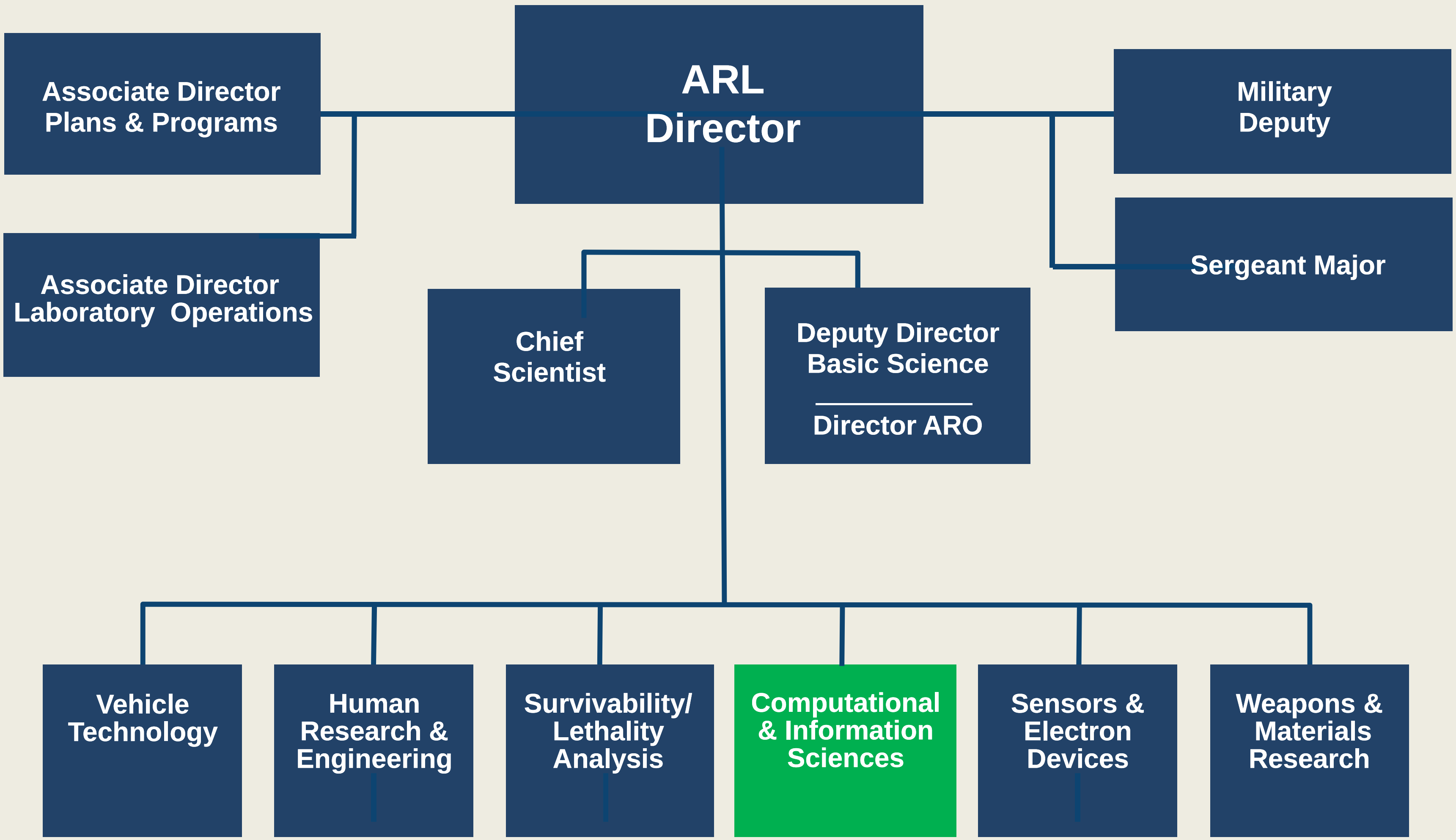
The Nation's Premier Laboratory for Land Forces.

Mission

DISCOVER, INNOVATE, and TRANSITION Science and Technology to ensure dominant strategic land power

Making today's Army and the next Army obsolete

Who We Are


ARL Director
Associate Director Plans & Programs
Associate Director Laboratory Operations
Military Deputy
Sergeant Major
Chief Scientist
Deputy Director Basic Science
Director ARO
Vehicle Technology
Human Research & Engineering
Survivability/ Lethality Analysis
Computational & Information Sciences
Sensors & Electron Devices
Weapons & Materials Research

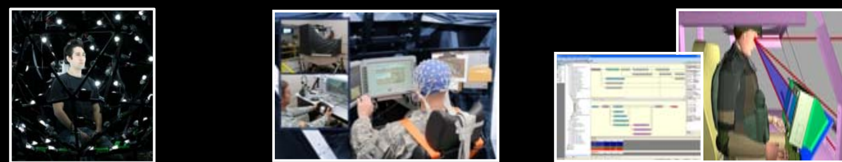
U.S. ARMY RDECOM

ARL S&T Campaigns

ARL

Extramural Basic Research
Steering and oversight of the systematic study to increase fundamental knowledge and understanding in physical, engineering, environmental, and life sciences related to long-term national security needs.

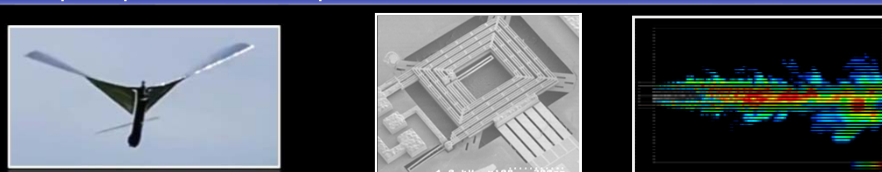
Human Sciences
Fundamental understanding of Warfighter performance enhancement, training aids, and man-machine integration..



Information Sciences
Fundamental understanding of information generation, collection, assurance, distribution, and exploitation.

Sciences for Lethality & Protection
Fundamental understanding of emerging technologies that support weapon systems, protection systems, and injury mechanisms affecting the Warfighter

Sciences for Maneuver
Fundamental understanding of the design, integration, control, and exploitation of highly adaptive platforms in complex environments



Computational Sciences
Fundamental understanding of computer hardware, high efficiency algorithms, and novel mathematical methods.

Materials Research
Fundamental understanding of structural, electronic, photonic, and energy materials & devices.

Assessment and Analysis
Quantitatively Assess the development and application of analytical tools and methodologies to quantitatively assess the military utility of Army, DoD, and select foreign combat systems.

Computational Sciences Taxonomy

Advanced Computing Architectures

- Emergent Architectures
- Tactical Computing
- Next Generation Computing Systems
- High Performance Networking and Memory

Computing Sciences

- Programming Environments
- Programming Languages
- Software Integration

BIG

High Performance Computing

DATA

Predictive Simulation Sciences

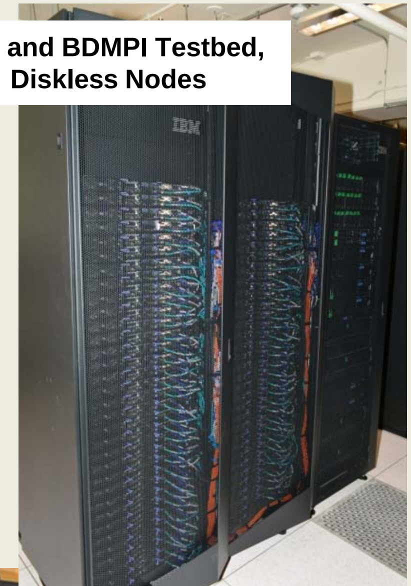
- Computational Math & Algorithms
- Scientific Computing
- Verification, Validation & Uncertainty Quantification
- Applied Computer Modeling and Analysis

Data Intensive Sciences

- Sciences of Large Data
- Computational Math for Data Analytics
- Real-time Data Access & Analytics

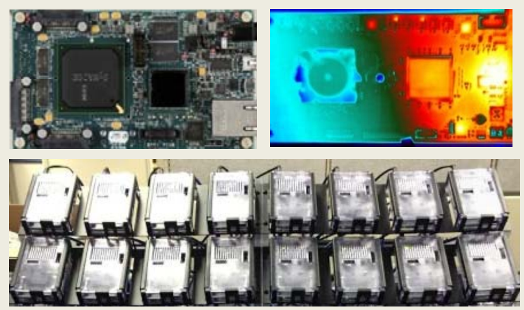
Our Hardware

Cray XC-40
Processing cores: 101,312
GPGPU's: 32 NVIDIA Tesla K40
Memory: 400 Terabytes
122 Terabytes: Solid State Disk or 'flash' storage
Peak Capability: 3.7 Petaflops



MPI and BDMPI Testbed, Diskless Nodes

Data Analytics Platforms

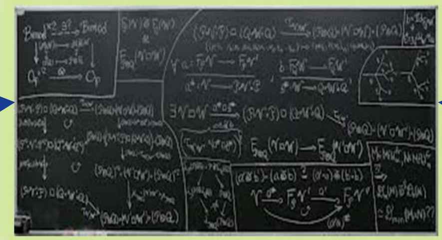


Neuromorphic system

48 nodes Intel Phi, 16 nodes Nvidia GPUs

Scientific Discovery

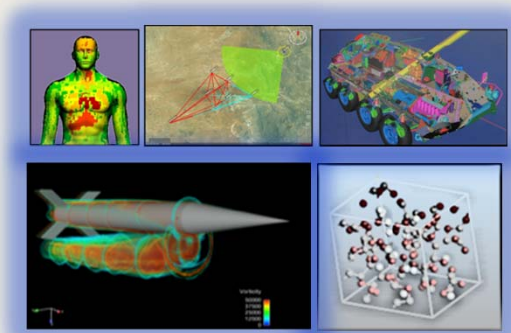
Theory

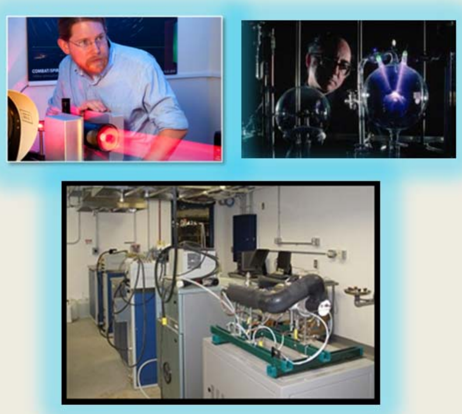


Theory embodied in computation

Hypotheses tested through experiment

SCIENTIFC METHODS





Computation complements experiment

Computation ⟷ Experiment

Data-Driven Discovery

- Modern scientific instruments are capable of collecting a petabyte or more of data per day
- HPC enables researchers to generate terabytes of data per case from physics-based computations for design, analysis, and discovery
- Developmental and operational tests, and training exercises, can generate a terabyte per day of data for events lasting multiple weeks from geographically separated areas and highly disparate data sources

Compounding the issue is the observation

- Between 2000 and 2010, disk storage capacity grew by a factor of *1,000* while disc access time improved by only a factor of *2*

Definitions

Big Data Characterized by

- *Volume*
  - How much data
- *Velocity*
  - The speed at which data arrives and the speed with which decisions based on it must be made
- *Variety*
  - Heterogeneity of storage platforms, data types, representation, semantic interpretation, and security classification or other distribution limitations
- *Veracity*
  - How trustworthy is the data, what is its uncertainty, and what is the error associated with it
- *Value*
  - What is the data worth

Background

Big Data = Data Science + Big Data Computing

- *Data Science*
  - The systematic study of digital data using scientific techniques
  - Observation, theory development, systematic analysis, hypothesis testing, and rigorous validation

- *Big Data Computing*
  - Novel and highly scalable computing paradigms
  - Dynamically provisioned networks
  - Innovative methods for processing and analysis
  - Domain-specific, value-driven applications

Changing Perspective

- Big Data is Big Business
- Big Data is Big Potential
- Big Data is Growing Exponentially
  - — Infeasible to store some data collections in their entirety
  - — Signal processing and data modeling must be used to extract manageable data samples
  - — Full data collection capability not used in all cases because of Volume
- We can no longer blithely collect or generate data and worry about its use and disposition after the fact
- We must recognize that like a materiel system, data has a *Life Cycle*, and we must plan for that Life Cycle

Example

Air Force Acquisition

- F-35 Program
  - Data from OEM, AF Program Office, AFRL, T&E, Training
  - 412th Test Wing has 455+ TB of data, Combined Test Force has more
    - Can collect 250K parameters at up to 10^6 Hz
- Navy and Marines have their own variants and data
- Foreign military sales add other partners
- Aircraft and program life cycle well defined; data life cycle is not
- The next airplane program will face the same issues again, only with even greater data generation and collection capability

Example

Army Test & Evaluation

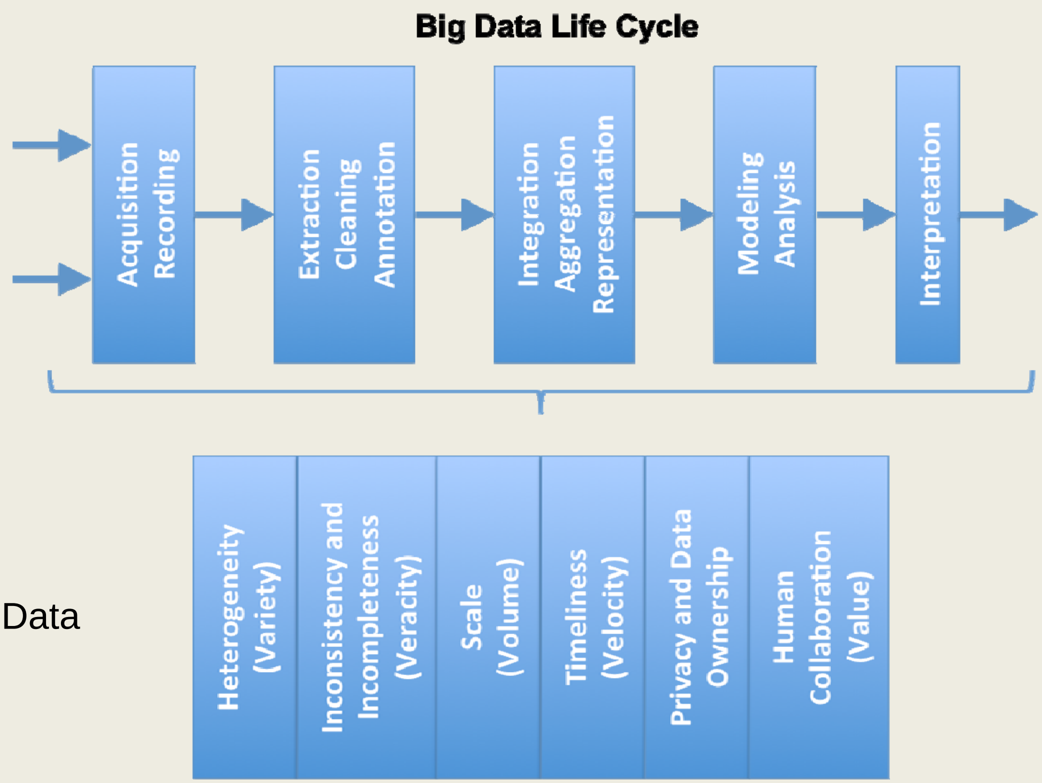
- Instrumented hundreds of vehicles in theater for data collection
- For the first time have the ability to understand if vehicles are being used as designed, on terrain for which they were designed
- Capturing braking, roll-over, vibration, crew compartment temperature, and other data in real time
- Have developmental test data for comparison and for forensics
- Have algorithms to reduce and display the data
- Need computational capacity (parallel algorithms and HPC platforms) to perform the analysis time-critically
- Need data framework to allow rapid access, comparison, analysis, display

Big Data Life Cycle

- Systems Engineering Approach to Big Data
- Requirements Based
- Grounded in how data will be used and who will use it
- Complements materiel system development life cycle

Life cycle stages (feedback can exist at every stage)

Cross-cutting challenges of Big Data



Big Data Requirements

Establish with Use Cases

- Involve all stakeholders
  - Acquisition
  - S&T
  - T&E
  - Training
  - OEMs
  - Etc.

- Identify common elements in requirements and embody those in standard or common tools and processes

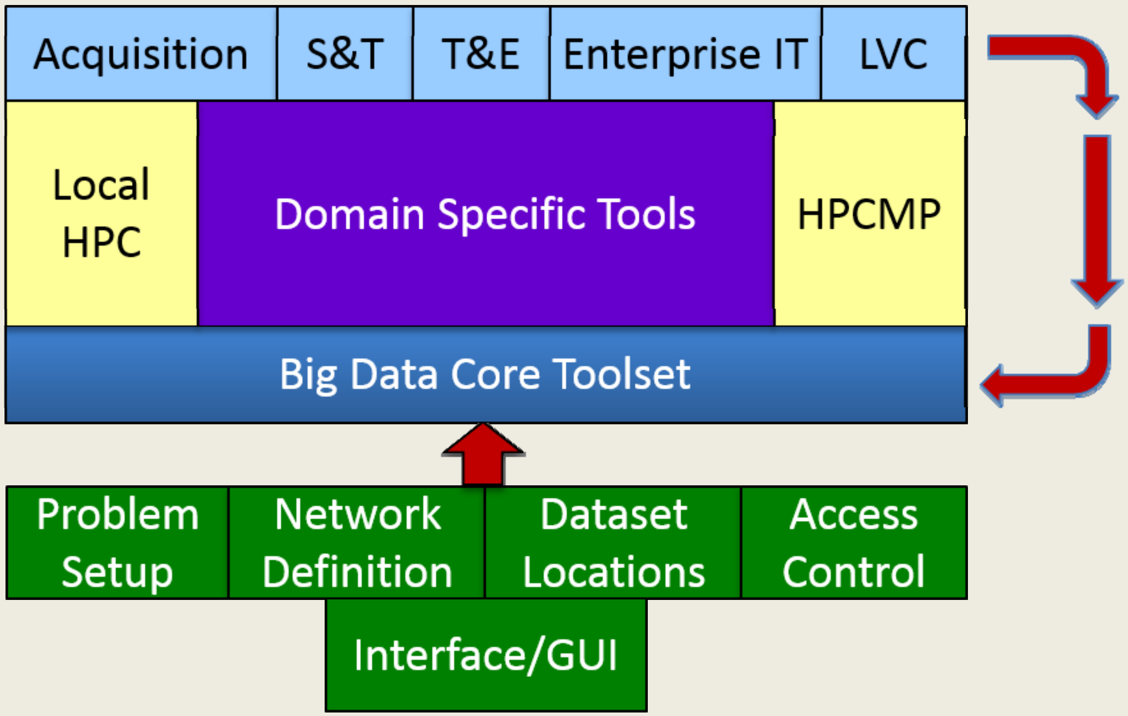
- Domain specific requirements met with domain unique tools

- Create flexible Big Data Framework to capture common and domain specific requirements

Big Data Framework

Extensible to other Stakeholders

- Standardize common tools and processes in Core Toolset
- Provide interface for domain specific tools
- Local and distributed computing, dynamically provision networks, control access to platforms, networks, tools, and data
- Use feedback to improve processes, tools, and implementation



Summary

- Big Data isn't new
  - We've always dealt with it
  - Now we must do so systematically and repeatably – it's growing exponentially
- The Army must remain a leader in timely processing of Big Data – the lives of our Warfighters depend on it
- We seek the best ideas from all sources, and invite collaboration for advancing data intensive science and computing

"There are more things in heaven and earth, Horatio, than are dreamt of in your philosophy." *Hamlet Act 1, scene 5*

Research Opportunities

- Neuromorphic Computing
- Deep Learning
- Tactical Cloudlets
- Software Defined Networks
- Highly Heterogeneous, Reconfigurable Systems
    - CPU, GPU, FPGA, Neuromorphic, D-Wave, etc.
- Advanced Graph Analysis
- Domain Specific Languages
- Stream Computing
- Highly Scalable Software and Systems

Computational Science Ecosystem

DOD HPC
DEPARTMENT OF DEFENSE
HIGH PERFORMANCE COMPUTING
MODERNIZATION PROGRAM

The Institute for Collaborative Biotechnologies
Biological Inspiration. Technological Innovation.

Extramural programs

ARL Campaigns & Unique facilities

Open Campus

Intern programs

Materials Research

Information Science

Sciences for Lethality and Protection

Sciences for Maneuver

ARL Advanced Computing and Networking Research Testbed

- Summer Institutes (high school)
- Internships (college)

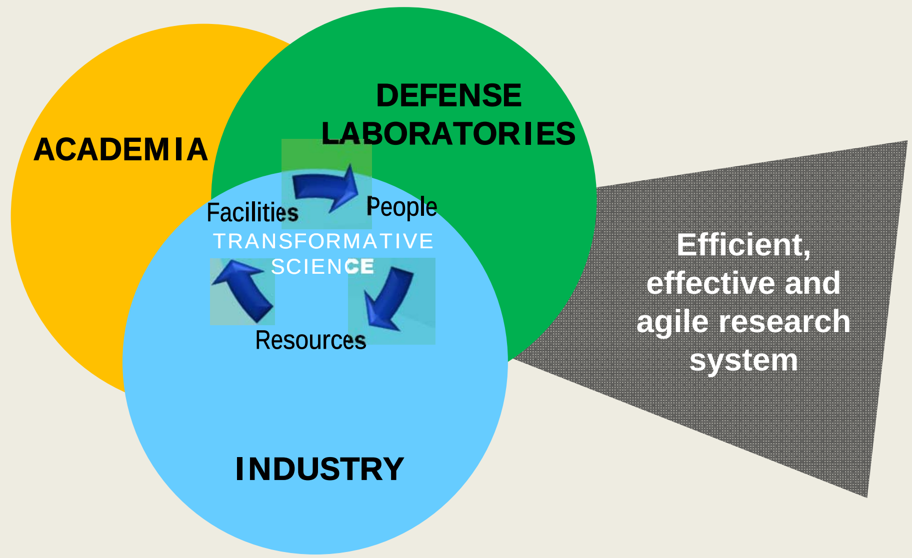
Piloting a New Laboratory Business Model

ATTRACT AND RETAIN BEST & BRIGHTEST

OPEN CAMPUSES

SHARED MODERN FACILITIES

INNOVATION PRACTICES



"We will need new technology over the next 10 years to make a leaner and more capable Army."

GEN Raymond T. Odierno
38th Chief of Staff, Army

Responding to the National Security Challenges of the 21st Century

Open Campus Website: http://www.arl.army.mil/opencampus/