DARPA and Data: A Portfolio Overview

William C. Regli
Special Assistant to the Director
Defense Advanced Research Projects Agency

Fall 2017

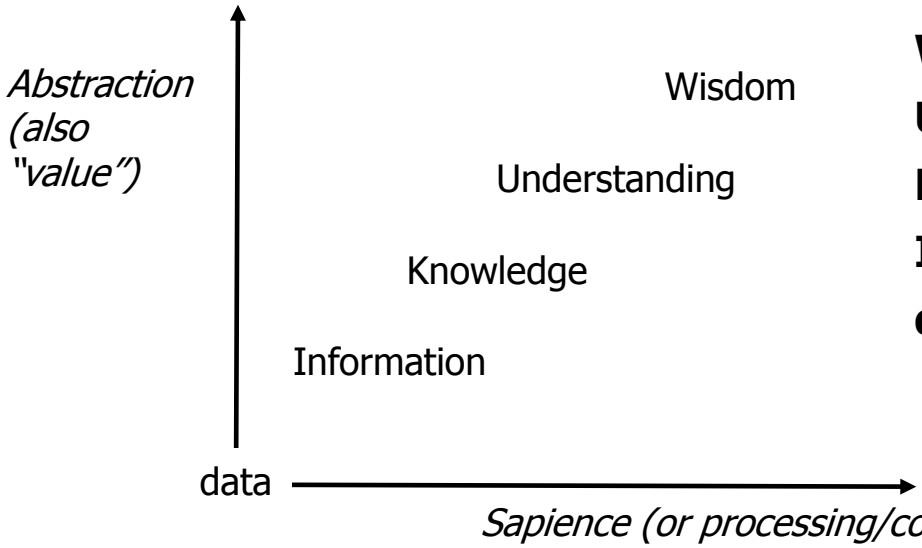
Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

DARPA Dreams of Data

- Investments over the past decade span multiple DARPA Offices and PMs
  - Information Innovation (I2O): Software Systems, AI, Data Analytics
  - Defense Sciences (DSO): Domain-driven problems (chemistry, social science, materials science, engineering design)
  - Microsystems Technology (MTO): New hardware to support these processes (neuromorphic processor, graph processor, learning systems)
- Products include DARPA Program testbeds, data and software
  - The DARPA Open Catalog
  - Testbeds include those in big data, cyber-defense, engineering design, synthetic bio, machine reading, among others
- Multiple layers and qualities of data are important
  - Important for reproducibility; important as fuel for future DARPA programs
  - Beyond public data to include "raw" data, process/workflow data
- Data does not need to be organized to be useful or valuable
  - Software tools are getting better exponentially, "raw" data can be processed
  - Changing the economics (Forensic Data Curation)
- Its about optimizing allocation of attention in human-machine teams

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

Working toward Wisdom



Wisdom: sound judgment - governance

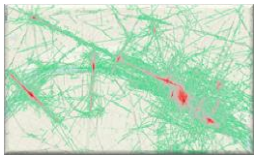
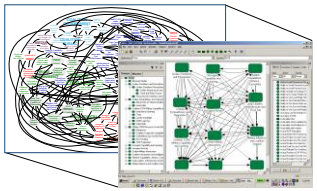
Understanding: explanation – why

Knowledge: correlation and composition – how

Information: facts - what, who, where, when

data: representing signals or stimuli

Example of the Data-Wisdom process for model evolution and application:

data	Information	Knowledge	Understanding	Wisdom
Data (signals) from multiple sources	Descriptive model (symbols, syntax)	 Correlative/predictive model (rules, processes, semantics)	 Explanatory model (abstractions, theories, etc)	Judgement, value, insight, instinct, principles, etc

DARPA Investments in the Data Ecosystem

Algorithms & Analytics

Services & Middleware

Information

Determine facts (what, who, where and when)

- LORELEI (I2O)
- MoDyl – Dynamics from data sets (DSO)
- RATS (I2O)
- Visual Media Reasoning (I2O)
- Global Autonomous Language Exploitation*

- Brandeis – Data privacy (I2O)
- EQUIPS – Uncertainty analysis tools (DSO)
- Memex – Web data/info search (I2O)
- SAFER (I2O)
- Simplex - Unified math frameworks (DSO)
- XDATA (I2O) →
- PROCEED*
- Signal processing programs*

Knowledge

Correlate and compose (networks and systems)

- CASCADE – Integ data/tools for robust SoS (DSO)
- CRAFT – Workflow/CAD tools (MTO)
- Data Driven Discovery of Models (I2O)
- Deep Extraction from Text (I2O)
- Insight (I2O)
- Make-It – Synthetic chemistry (DSO)
- MENTOR2 – CAD data/tools (DSO)
- Modelling Adversarial Activity (I2O)
- Network Defense (I2O)
- Next Gen Soc Sci – Data/tools (DSO)
- QCR (I2O)
- Sigma – System for CBRNE detection (DSO)
- TRADES - Eng design env (DSO)
- Personalized Assistant that Learns*

- Cortical Processor (MTO)
- EdgeCT (I2O)
- GRAPHS – Graph analysis tools (DSO)
- HIVE – Graph problem HW/MW (MTO)
- Media Forensics (I2O)
- MUSE (I2O)
- SafeWare (I2O)
- Deep Learning*

Understanding

Explain (causal mechanisms)

- AIDA (I2O)
- Big Mechanism (I2O)
- Biochronicity – Tools gene regulatory nets (DSO)
- Causal Exploration (I2O)
- MATRIX– Multi-scale matls data/models (DSO)
- Mine Better, Fund Faster – Surprise ID (DSO)
- SocialSim (I2O)
- Synergistic Discovery and Design (I2O)

- BRASS (I2O)
- Explainable AI (I2O)
- FunLOL – Framework for AI/ML (DSO)
- MSEE – Tools for machine understanding (DSO)
- PPAML (I2O)

Algorithms & Analytics

Services & Middleware

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

*completed program

DARPA Investments in the Data Ecosystem

Algorithms & Analytics

Services & Middleware

Information

Determine facts (what, who, where and when)

LORELEI (I2O)
MoDyl – Dynamics from data sets (DSO)
RATS (I2O)
Visual Media Reasoning (I2O)
Global Autonomous Language Exploitation*

Brandeis – Data privacy (I2O)
EQUIPS – Uncertainty analysis tools (DSO)
Memex – Web data/info search (I2O)
SAFER (I2O)
Simplex - Unified math frameworks (DSO)
XDATA (I2O)
PROCEED*
Signal processing programs*

Knowledge

Correlate and compose (networks and systems)

CASCADE – Integ data/tools for robust SoS (DSO)
CRAFT – Workflow/CAD tools (MTO)
Data Driven Discovery of Models (I2O)
Deep Extraction from Text (I2O)
Insight (I2O)
Make-It – Synthetic chemistry (DSO)
MENTOR2 – CAD data/tools (DSO)
Modelling Adversarial Activity (I2O)
Network Defense (I2O)
Next Gen Soc Sci – Data/tools (DSO)
QCR (I2O)
Sigma – System for CBRNE detection (DSO)
TRADES - Eng design env (DSO)
Personalized Assistant that Learns*

Cortical Processor (MTO)
EdgeCT (I2O)
GRAPHS – Graph analysis tools (DSO)
HIVE – Graph problem HW/MW (MTO)
Media Forensics (I2O)
MUSE (I2O)
SafeWare (I2O)
Deep Learning*

Understanding

Explain (causal mechanisms)

AIDA (I2O)
Big Mechanism (I2O)
Biochronicity – Tools gene regulatory nets (DSO)
Causal Exploration (I2O)
MATRIX– Multi-scale matls data/models (DSO)
Mine Better, Fund Faster – Surprise ID (DSO)
SocialSim (I2O)
Synergistic Discovery and Design (I2O)

BRASS (I2O)
Explainable AI (I2O)
FunLOL – Framework for AI/ML (DSO)
MSEE – Tools for machine understanding (DSO)
PPAML (I2O)

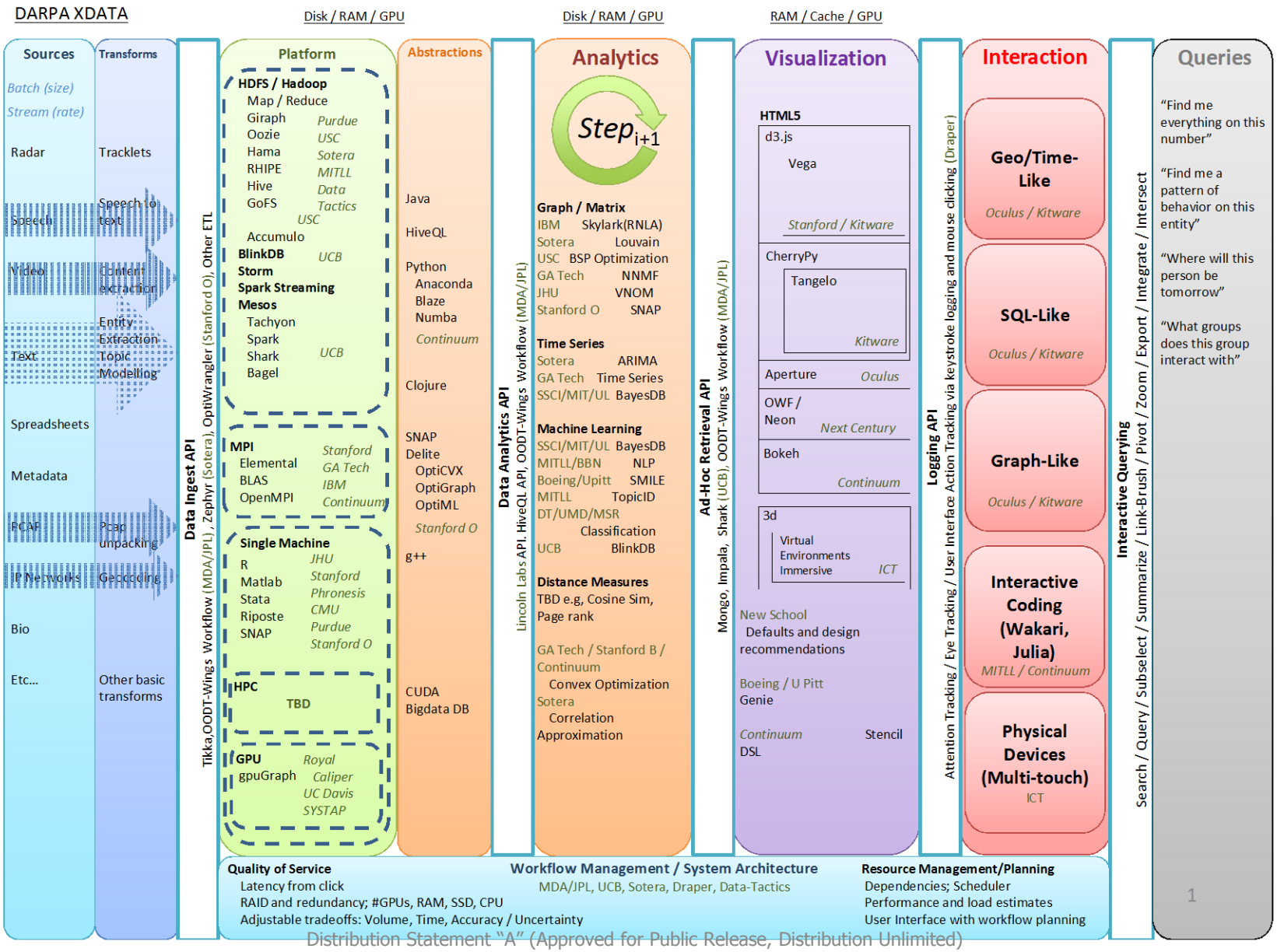
Algorithms & Analytics

Services & Middleware

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

*completed program

X-Data: Common Data Cyber-Infrastructure



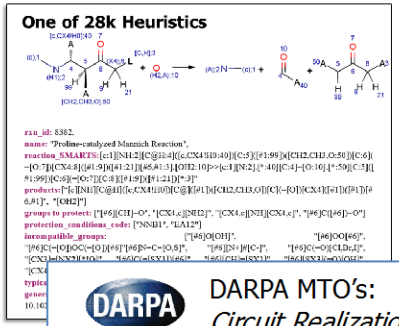
Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

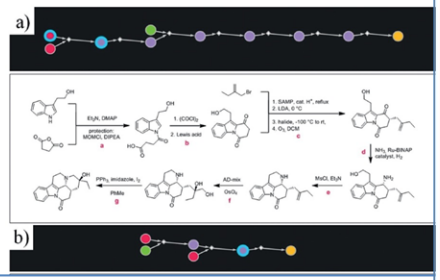
Examples: Human-Machine Symbiosis

DARPA DSO's MAKE-IT: Revolutionizing Synthetic Chemistry

Digitize Structure → **Generate Pathways** → **Score Pathways** (Chemical & Reaction Scoring Function (CSF & RSF)) → **Rank Output** (Estimate ease of implementation)

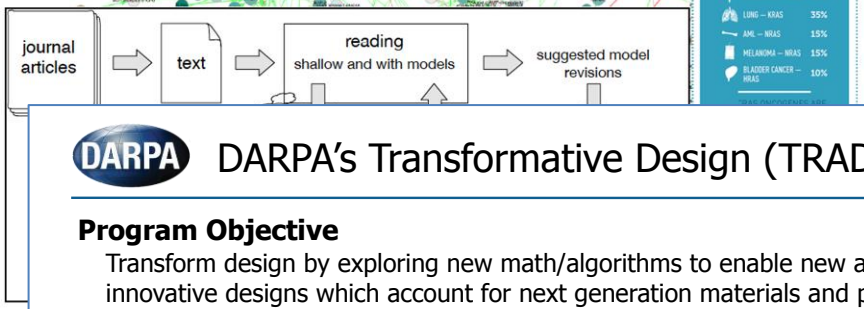
InChI=1S/C4H3N7O4/c12-10(13)5-3-1-2-4-8(3)6-7-9(4)11(14)15/h1-2,5H

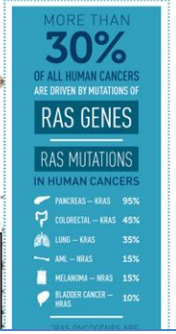




DARPA's Big Mechanism: Machine Reading vs Cancer

Goal: Machines that help humans to model and analyze very complicated systems by reading fragmented literatures and assembling and reasoning with models.





DARPA MTO's: *Circuit Realization at Faster Timescales (CRAFT)*

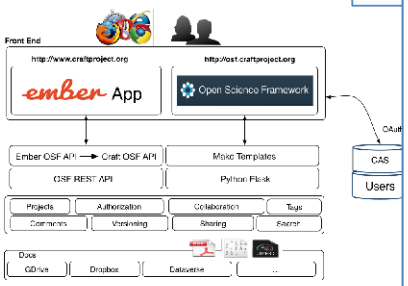
CRAFT aims to provide solutions to the three major obstacles restricting custom IC design and fabrication for DoD systems.

DESIGN

- Design requires 18-24 months of effort
- Design verification takes far too much effort
- Fab cycles are too long and too uncertain
- Access to leading-edge CMOS is difficult

PORT/ MIGRATE

- Designers are limited to one foundry
- Migration of designs from one node to another is diff… expensi…



REPOSITORY

- Severe lack of IP reusability for DoD designs
- Current audit model for custom IC design/hardware security is broken

CRAFT aims to establish a data location and methodology to ensure 50% IP* reuse by DoD performers

CRAFT 's goal is to enable more efficient custom IC design/fabrication to enable HIGH performance electronic solutions FASTER and with more FLEXIBILITY

* IP – Sub-circuits used for modern custom ICs

Approved for Public Release, Distribution Unlimited

44

DARPA's Transformative Design (TRADES)

Program Objective

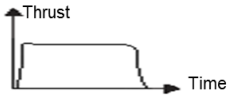
Transform design by exploring new math/algorithms to enable new and innovative designs which account for next generation materials and processes

Expected Outcomes

- Leverage computation and data at unprecedented scales
- Deal with significantly more complex and coupled design problems
- Lower the time and expertise required to create complex engineered products

Human

Describe the problem

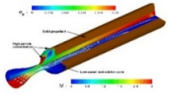


Show me the options

⇔

Computer

Generate shape & material



Find the best alternatives

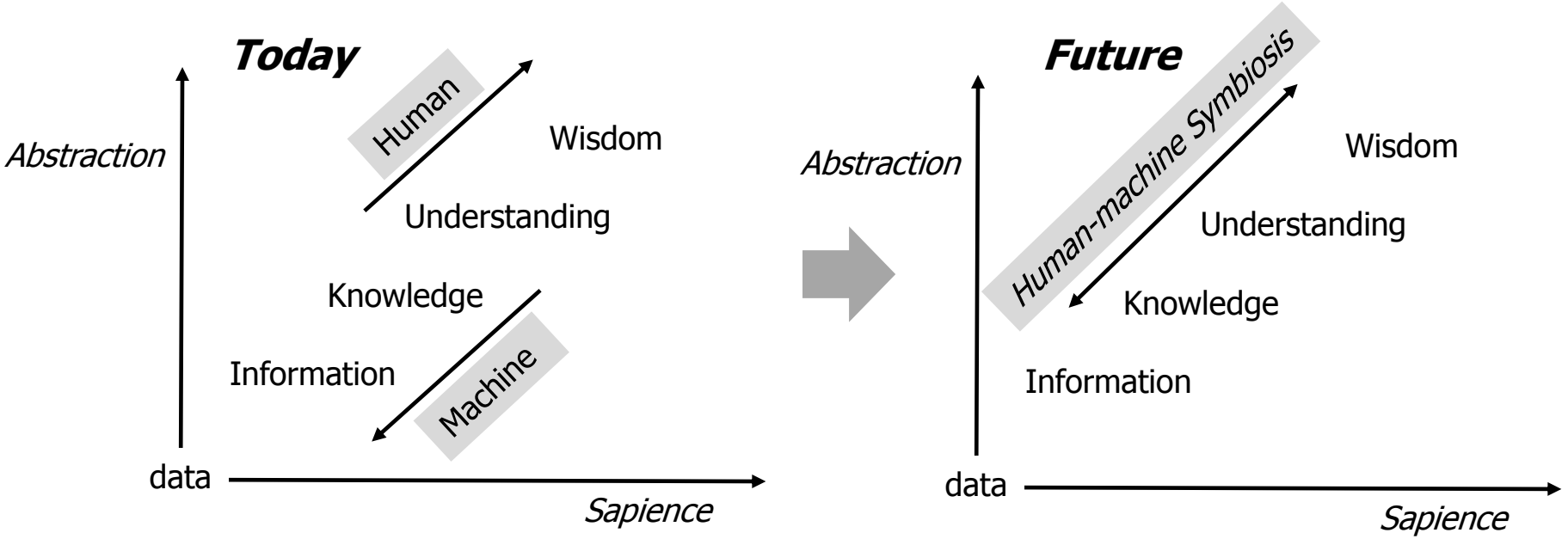
http://space.stackexchange.com/questions/12541/why-is-there-a-hole-in-solid-rocket-engines/12542

http://arrow.utias.utoronto.ca/~groth/Research.html

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

37

Summary



Unbound Computation & Data

Exploit computation and larger scales; leverage machine learning and AI;
Produce and harvest data in more disruptive ways! ← requires culture change

Advancing the software toolbox

An ecosystem of interoperable building blocks/tools for machine-augmented problem solving; make them easier to use ← a limiting issue

Problem-Process-People Co-design

Simultaneous redesign of problem and human-machine dynamic:
Thinking faster, better and functioning at higher levels ← culture change

www.darpa.mil