Deep Data Analytics in Support of Acquisitions and Tradespace Analysis

Engineered Resilient Systems Track
NDIA Systems and Mission Engineering Conference 2018
October 24, 2018

David Stuart
Associate Technical Director
Information Technology Laboratory

ITL INFORMATION TECHNOLOGY LABORATORY

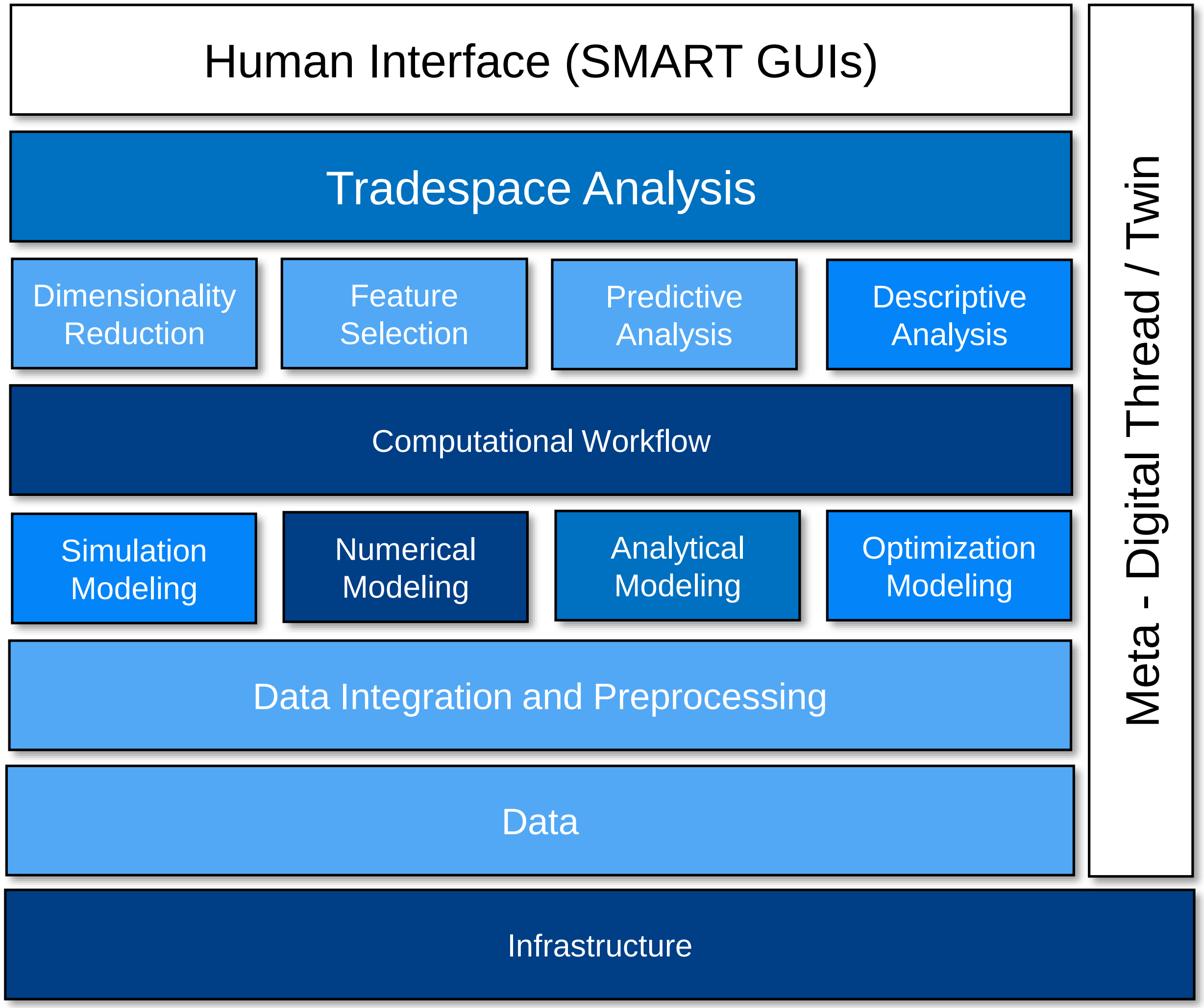
ERS ENGINEERED RESILIENT SYSTEMS DEPARTMENT OF DEFENSE

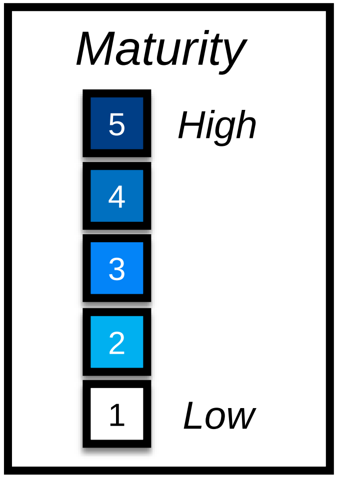
Distribution A: Approved for public release

Topics

- Current State of Tradespace Analysis
- New thinking
- Technology
- Data analytics ecosystem and processes
- A.I.
- Machine Assisted Tradespace Analysis

Current State





New Thinking (10x)

Improve decision making through the integration of advance computing into the decision-making process. Humans perform higher-level strategic thinking, while machines conduct lower-level decision

Today

- **Marginalize**
  - Analysis is severely restricted
  - Machines play a secondary role to humans
- **Limited**
  - Analytics do not scale to large large problems
  - Analysis is very swallow
- **Brittle**
  - Single point solutions

Change Drivers

- ***Growth in information requires machines to take a more activity participate in decision making***
- ***Humans will conduct high-level decisioning …machines work to make lower-level decisioning***
- ***Data sizes will overwhelm decision-makers and complicate the decision making process***
- ***Deep Analytics - breadth and depth of the analysis help service insights from all types of data***
- ***Capable of operating on data sets at the petabyte scale***
- ***Prioritize important points for analysis by humans***
- ***decisions will be subdivided into levels machine-level and human-level***

Tomorrow

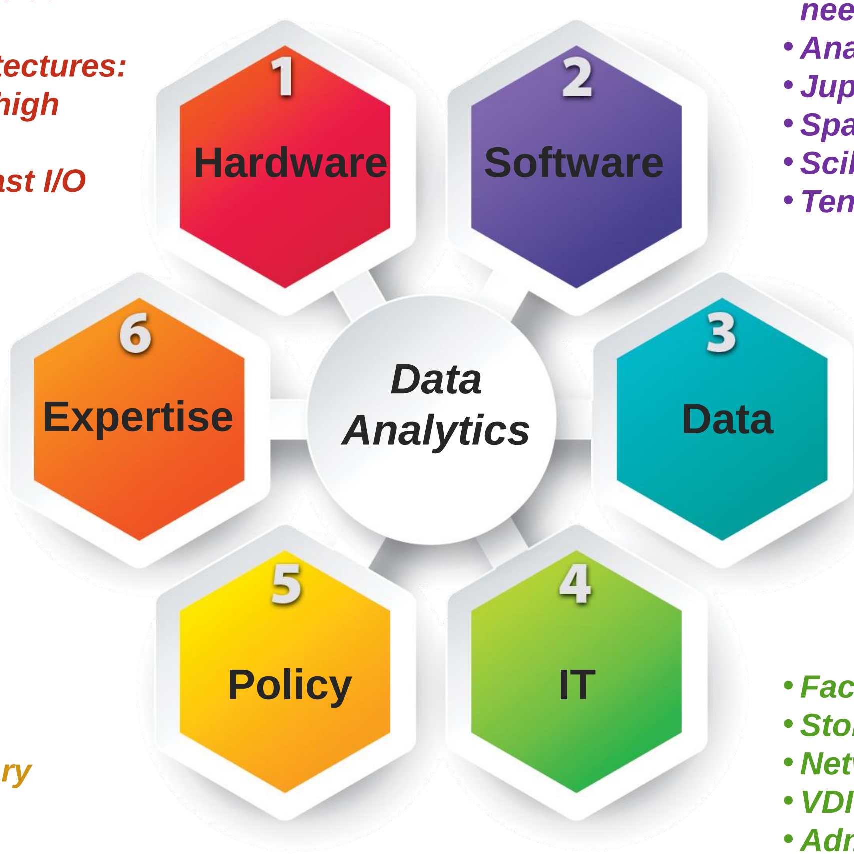
- **Go Faster**
  - Host analysis on HPC
- **Think Deeper**
  - Methods scale to address large complex problem spaces
  - Inclusion of a breadth of information
  - Data and knowledge are integrated
- **Be Resilient**
  - Identify a set of alternatives as opposed to a single solution

Competencies

- *Machines that can address our largest problems*
- *Blended computing architectures:*
  - *Numerical - distributed, high speed interconnects*
  - *Data - shared memory, fast I/O*

- *Computational and data scientists assist in problem step up, execution, and visualization*

- *Integrate necessary policies*



- *Leverages open source capabilities*
- *Python ,R, and C/C++ (when needed)*
- *Anaconda - package management*
- *Jupyter Notebooks*
- *Spark, Galaxy, Dakota*
- *Scikit Learn (machine learning)*
- *TensorFlow (deep learning)*

- *Terabytes collected - Unorganized and inaccessible*
- *Streamline data wrangling*
- *Minimize the movement of data*
- *Leverage database technologies*
  - *SQL and noSQL*

- *Facilities*
- *Storage (hot, warm, and cold)*
- *Networks (10G+)*
- *VDI*
- *Administration of machines*
- *Security - (monitoring, patching, etc.)*

Data Analytics Ecosystem

Mass Storage Services

- ERDC's MSAS System (Gold)
- 1PB SAMFS filesystem
- 25PB current on tape (80 PB capacity)
- 3 Oracle T54 servers running Solaris 11
- Data transfers are equally distributed among the 3 T54 servers)

Data Analytics

- Anaconda Enterprise
- Spark
- Machine Learning
- Deep Learning

High Speed Network (100G)

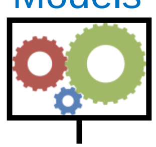
General Computing Services

- Virtualization
- Persistent Services
- Data Management
- Interactive Access

Computational Science

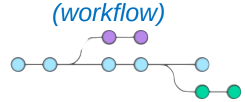
- 126,192 compute cores
- 544 Knights Landing nodes
- 32 GPGPUs
- 437 terabytes of memory
- **Adding Urika software stack**

Models



Dakota
(optimizer)

Galaxy
(workflow)

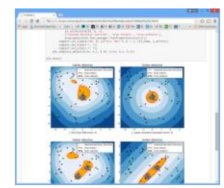


Decision Portal

Web connection

Decision Makers

Jupyter Notebooks



Web or VDI connection

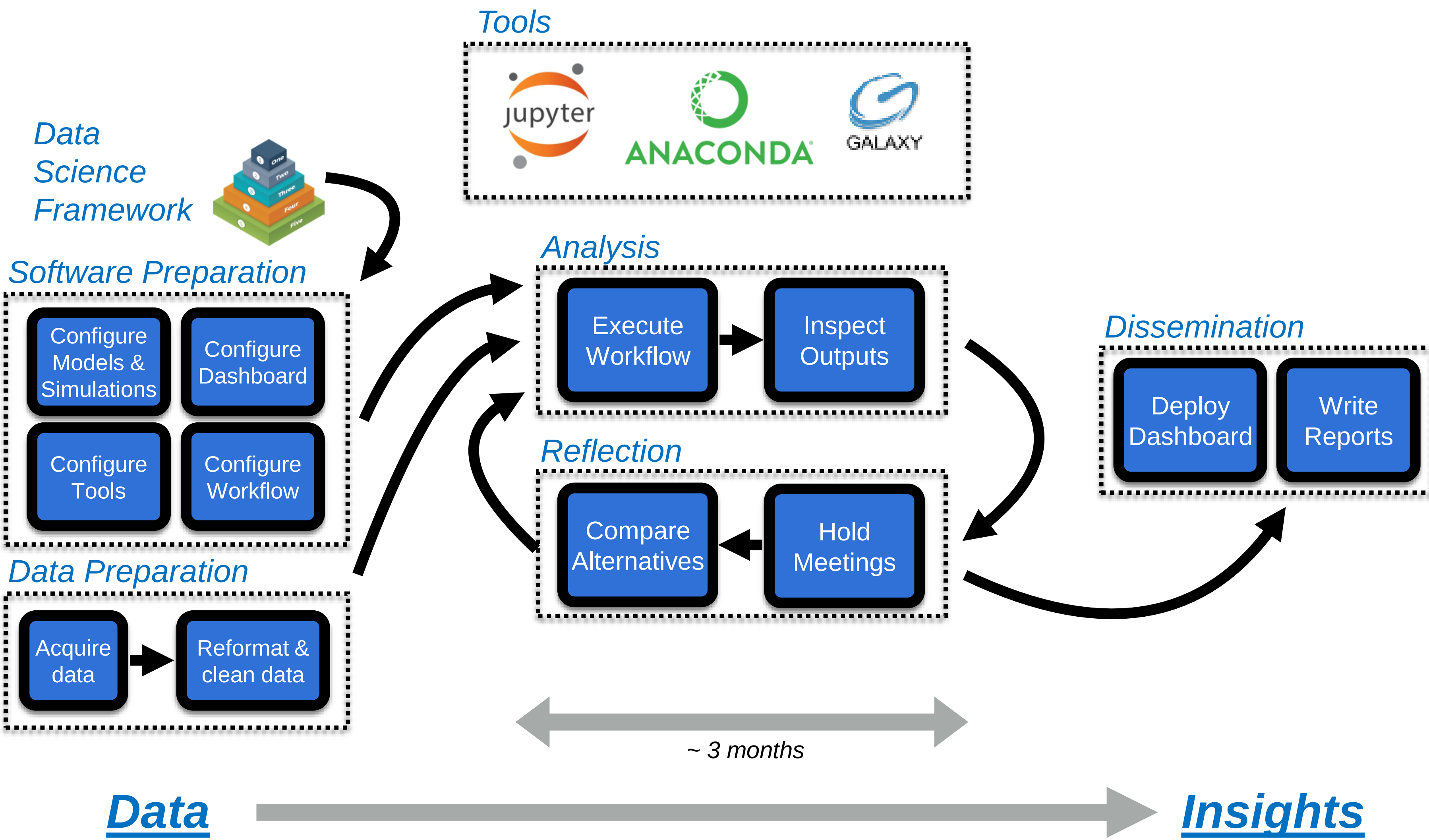
Data Scientists

NVIDIA DGX-1

Computational Science & IT ←

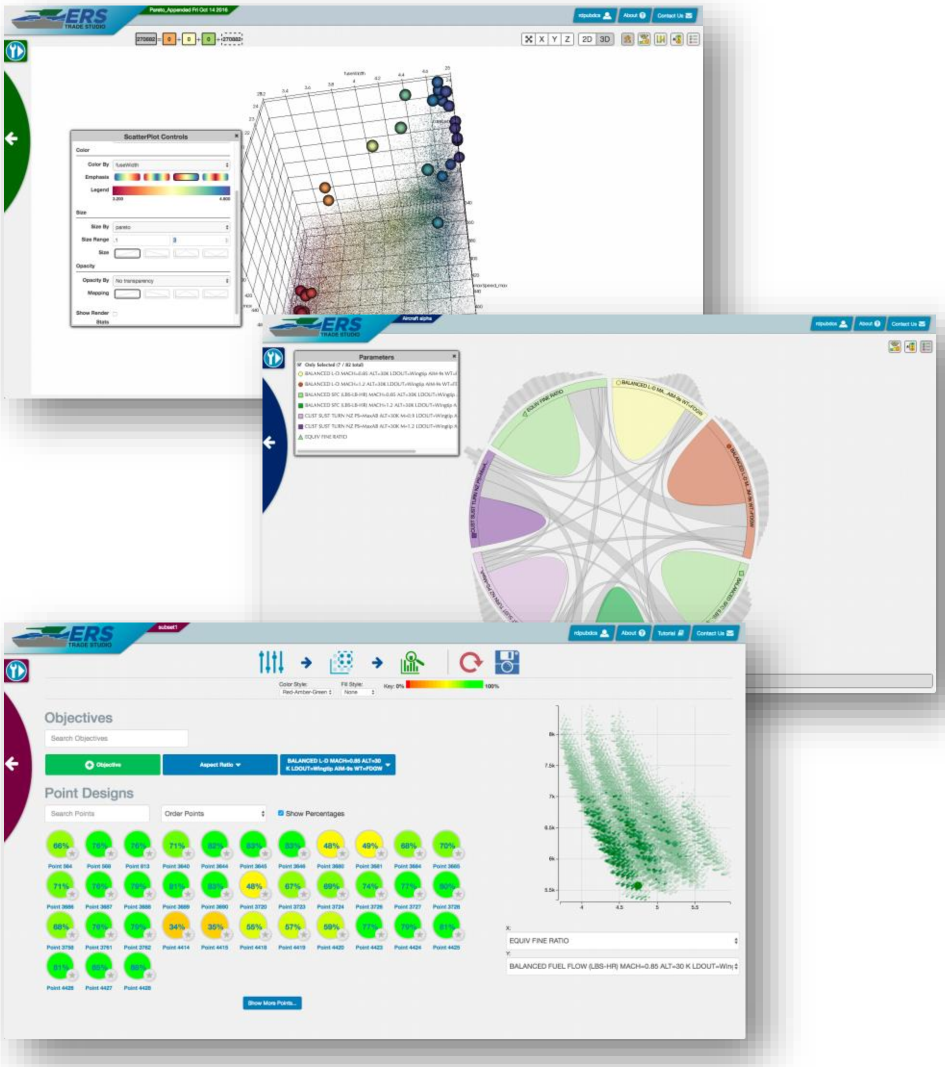
→ **Data Science & Decision Making**

Data Science Workflow



Current Tradespace Workflow

- Heavy Visualization
- Manual process
- Millions of designs considered, but only a few in detail

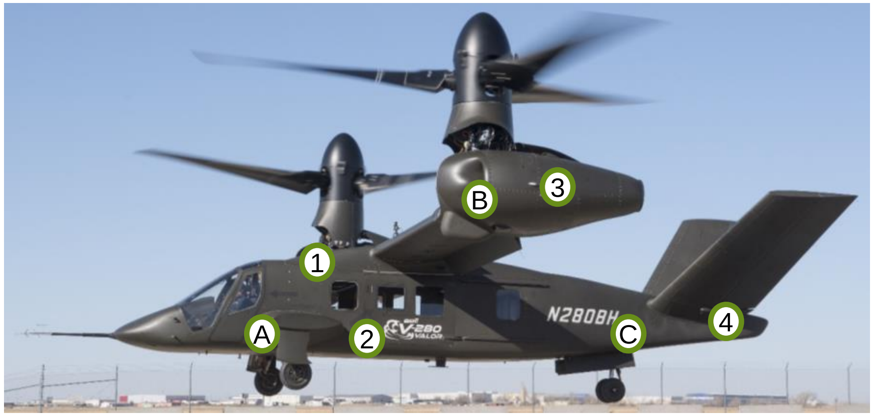


A.I.

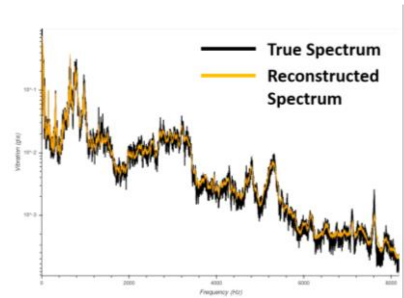
With data…

Without data…

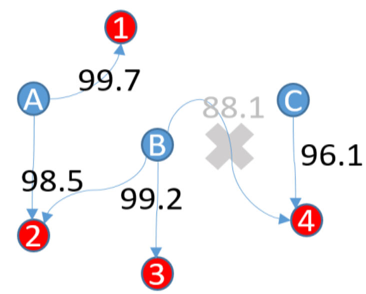
With Data - Virtual Sensors



Historical operational sensor data can be used to study how many sensors are needed and where they need to be placed.



If a sensor can be inferred from other sensors with a high degree of accuracy then instead of fielding a physical sensor, a “virtual sensor” model, developed on DSRC-HPC, can be used.



The minimum virtual sensor cover is the minimum set of physical sensors necessary to infer ALL sensors with some required level of accuracy

A **minimum virtual sensor cover** for FVL would save space, weight and power, extending range and lifting capacity and **save hundreds of millions of dollars** in up front manufacture and life cycle maintenance cost.

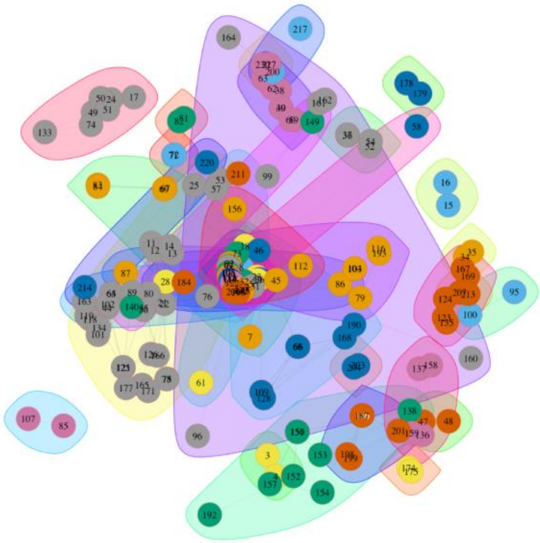
Without data – AlphaGo Zero

- Learn from scratch
- No historical data
- Data generated from unsupervised training
- 20 days of training to beat world champion

Complexity
Chess: 10^{120}
Go: 10^{174}

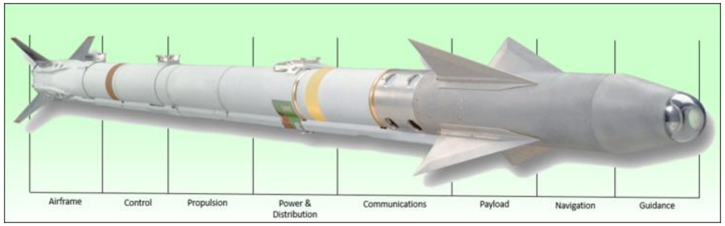
Machine Assisted Tradespace Analysis - Needs

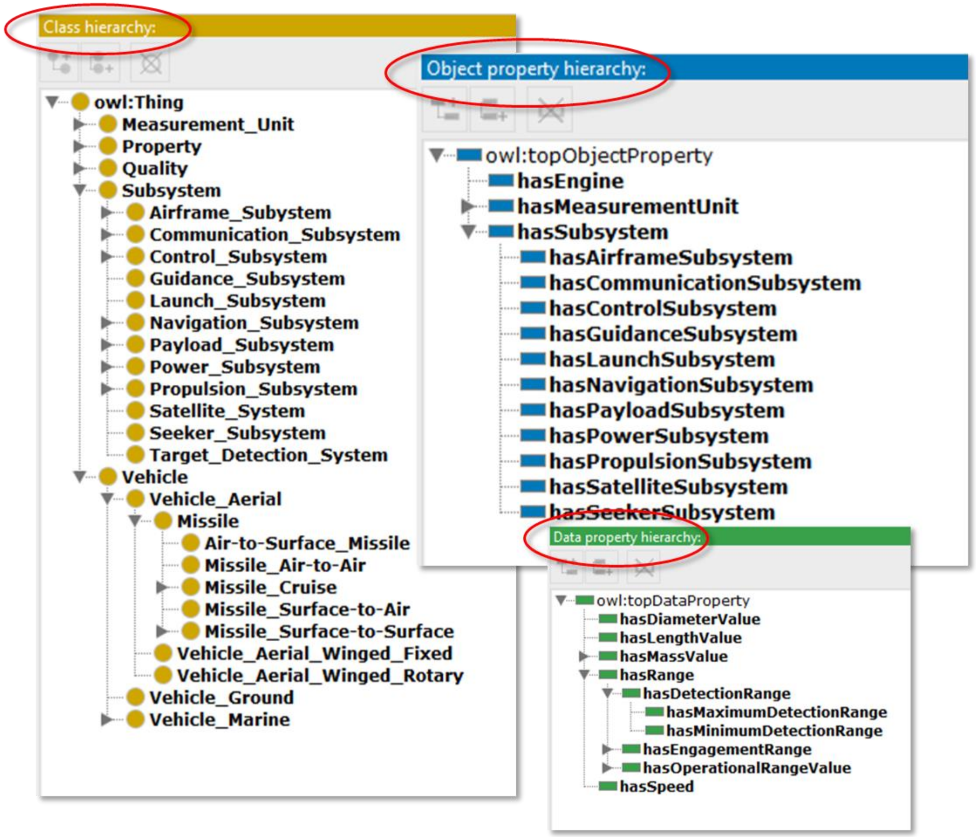
- Full definition of the problem
- Need win condition and rules
  - Capabilities
  - Constraints
- Functional Framework (Driver)
- No beginning tradespace



Ontologies

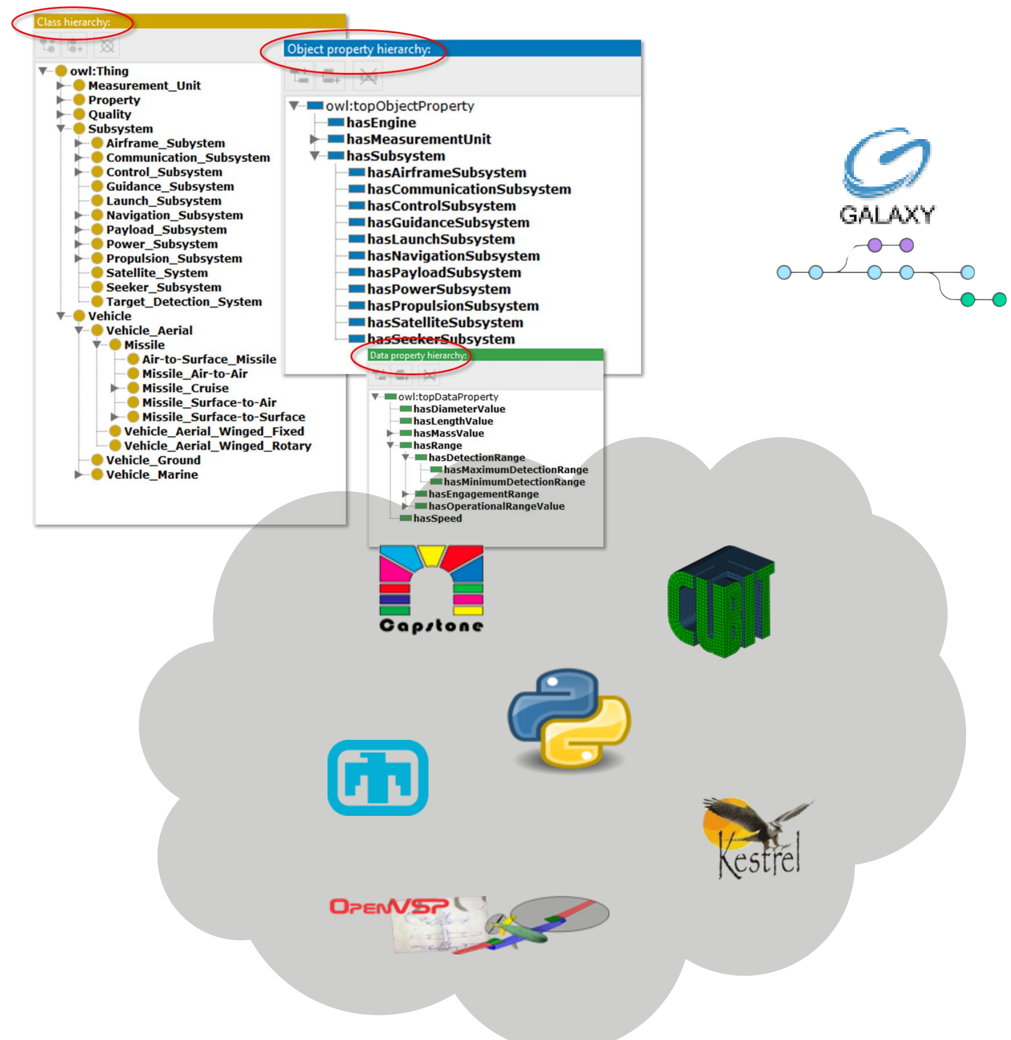
- Machine Assisted Driver
- Provides Structure
- Defines Semantics
  - User understanding
  - Machine understanding
- Defines constraints
- Drive digital twin



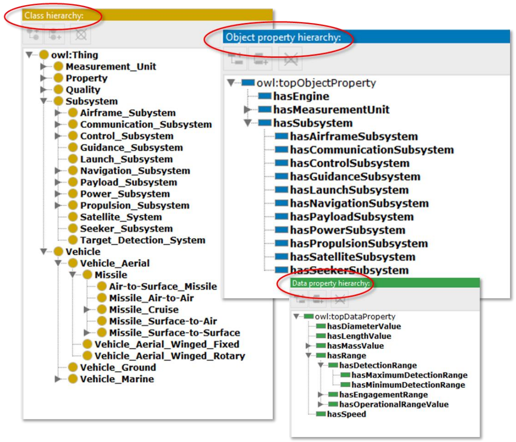


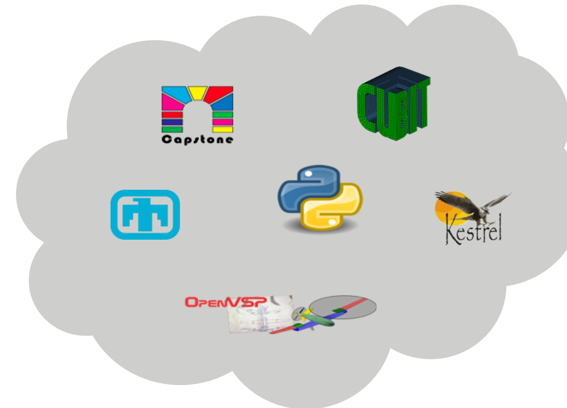
Map Ontologies to models

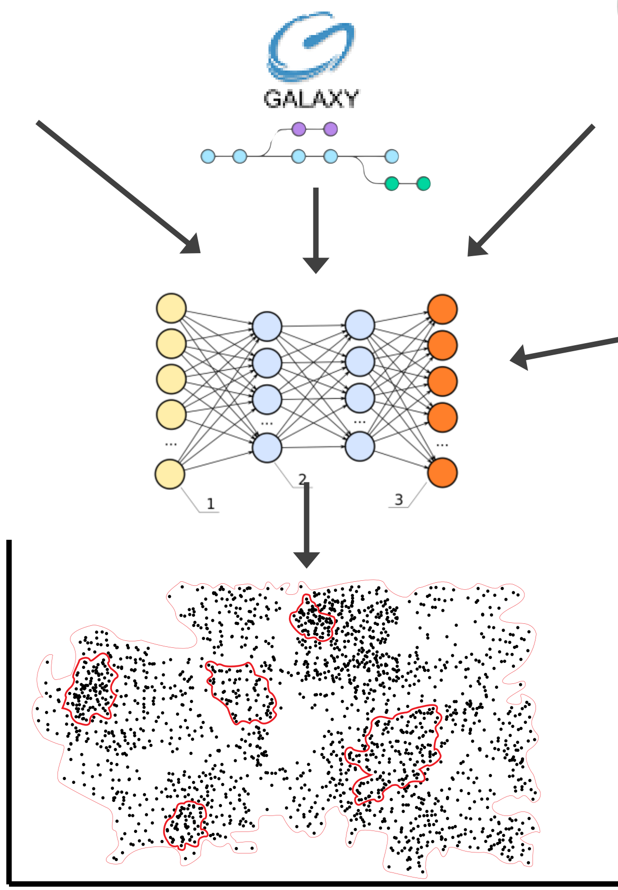
- Attributes within the ontology are mapped to inputs of the available models
- Performance metrics mapped to outputs



Machine Assisted tradespace analysis
How does it work





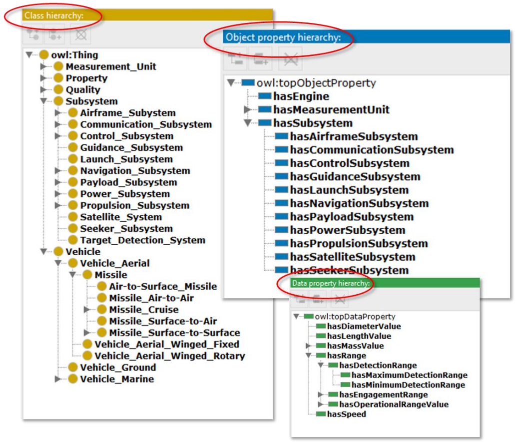


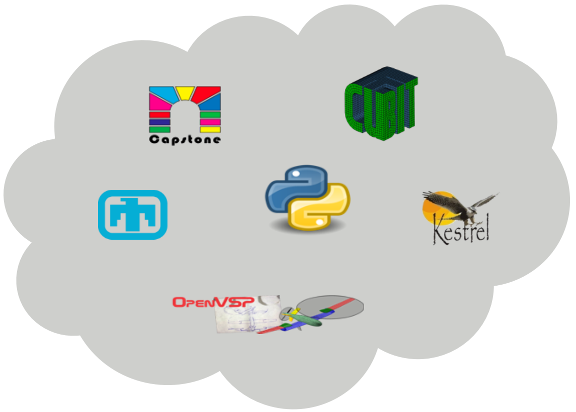
Win condition – desired capabilities

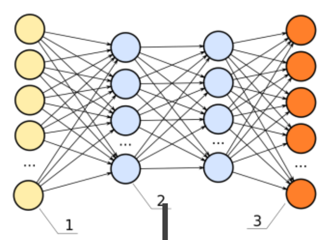
First Iteration

- Low fidelity
- Millions of designs
- Clusters of high performance designs
- Select 1 or more clusters

Machine Assisted tradespace analysis
How does it work

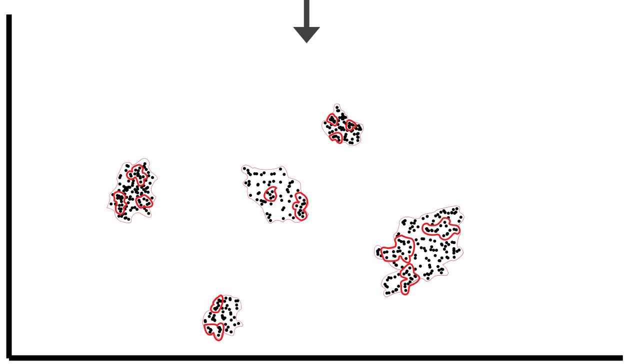




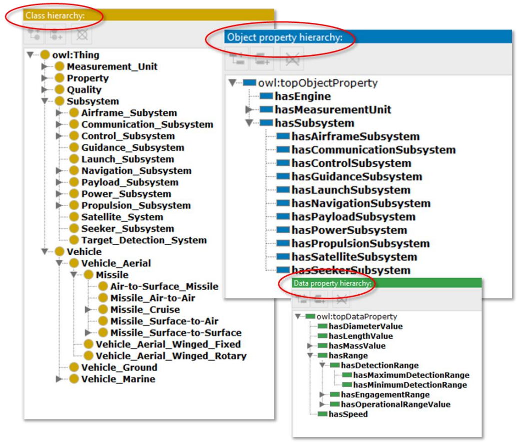


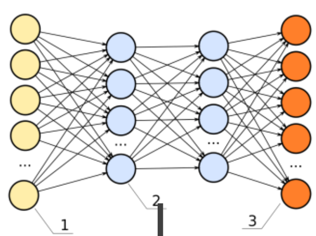
Win condition –
desired capabilities

Second Iteration
- Moderate fidelity
- 10,000’s of designs
- Clusters of high performance designs
- Select clusters within clusters



Machine Assisted tradespace analysis
How does it work

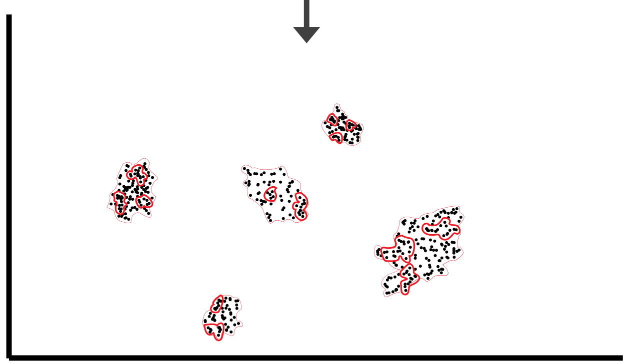




Win condition –
desired capabilities

Nth Iteration

- Increasingly higher fidelity
- Decreasing number of designs
- Clusters of high performance designs
- Select clusters within clusters



Final output

- Set(s) of designs
- Complex constraint sets

Contact

David Stuart
US Army Engineer Research and Development Center
Information Technology Laboratory
david.c.stuart@erdc.dren.mil
601-634-3174

Distribution A: Approved for public release