Data-to-Decisions S&T Priority Initiative

Dr. Carey Schwartz

PSC Lead

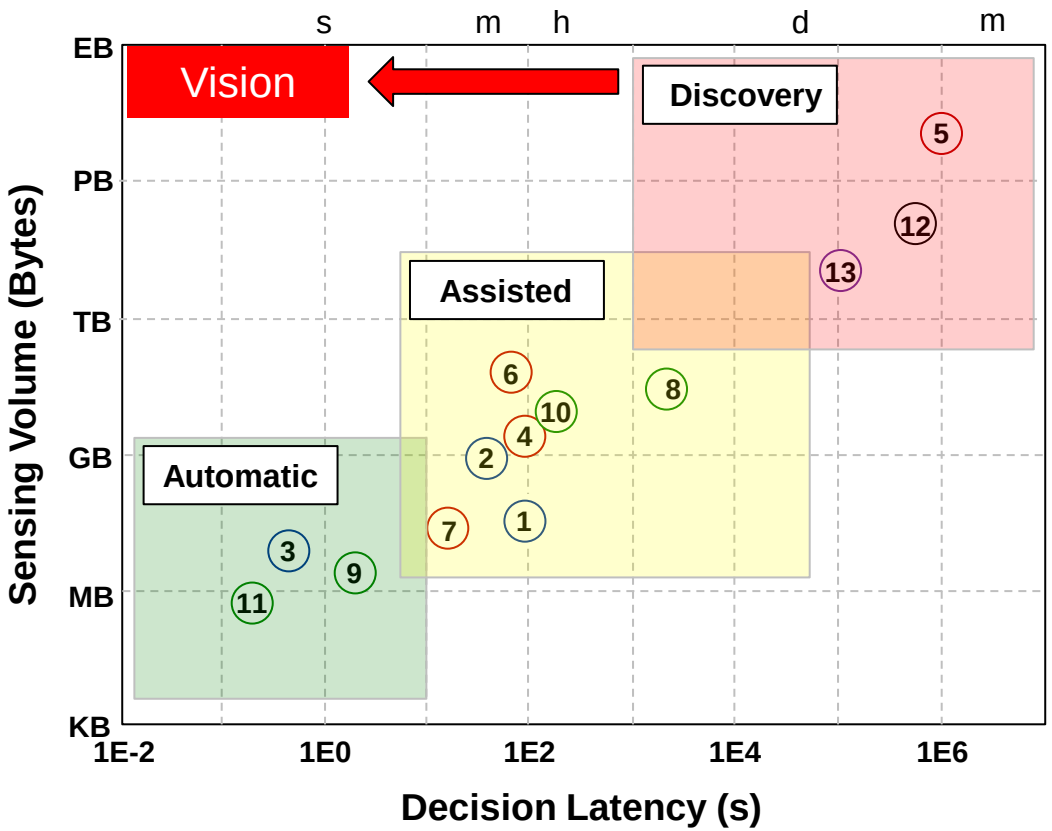
Office of Naval Research

NDIA Disruptive Technologies Conference

November 8-9, 2011

Washington, DC

Data-to-Decisions Systems Issues – Time and Volume



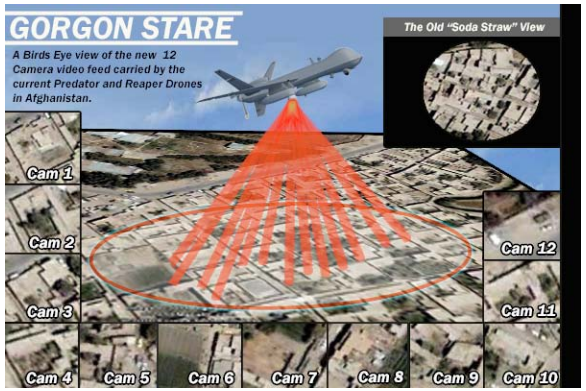
- **Defend United States**
  - 1. Containerized Nuclear Weapon*
  - 2. Blackmail ICBM*
  - 3. LACM off barge
- **Counterinsurgency**
  - 4. Unguided Battlefield Rocket
  - 5. Insurgencies
  - 6. IEDs
  - 7. Small fast attack craft
- **Anti-Access Environments**
  - 8. Quiet submarines
  - 9. MARV (Intercept)
  - 10. Mobile long-range SAMs
  - 11. Co-orbital ASAT
- **Security Capacity**
  - 12. Stability Operations
- **Counter WMD***
  - 13. Loose Nukes

National security decision systems span all QDR missions with a focus on finding threats in a specified data volume with limited manpower within a specified time window

Data-to-Decisions Systems Issues - Personnel

Predator Sensor

Increasing Resolution and Coverage



Analysts

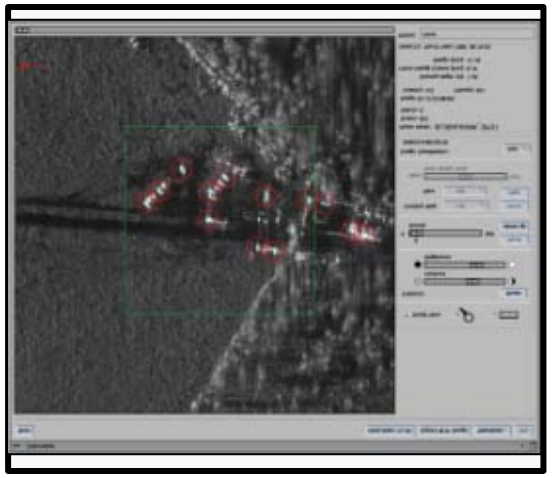
Number of Highly Skilled and Trained Analysts Remains Constant or Decreases

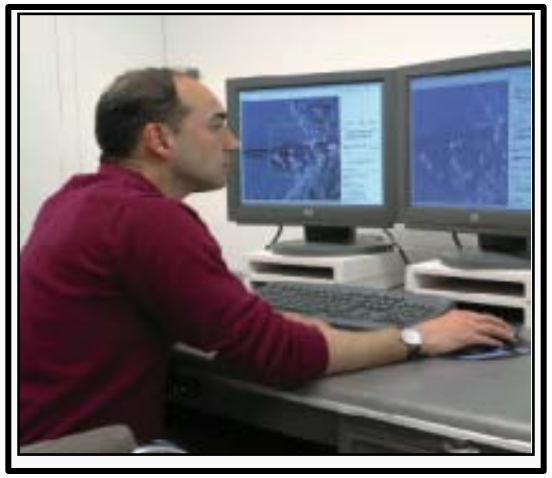
Analysts

National security decision systems span all QDR missions with a focus on finding threats in a specified data volume with limited manpower within a specified time window

D2D Technology Assessment

Data Management Layer	Analytics Layer	User Interface Layer
• Moderately Mature • Driven by IT Industry	• Immature • Driven by Defense	• Moderately Mature • Driven by IT Industry





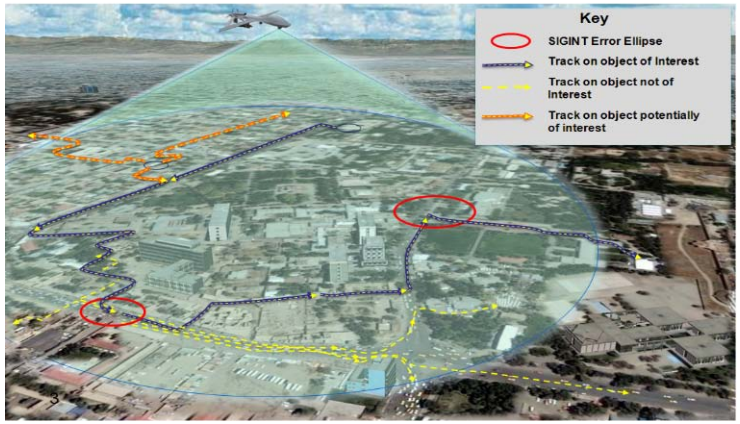
Current assessment is that unstructured data analytics is the most challenging and critical component of D2D

ASD D2D program intends to provide representative data of various types that have associated ground truth to support development and evaluation of algorithms and systems in a SOA to be made available

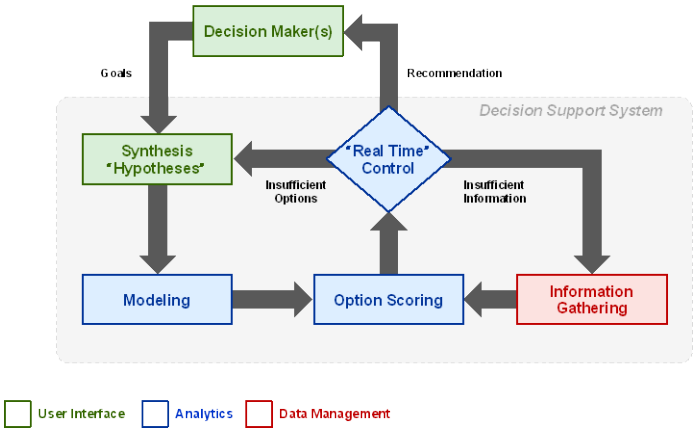
Challenge Problem and Framework for Analysis

Illustrative Challenge Problem

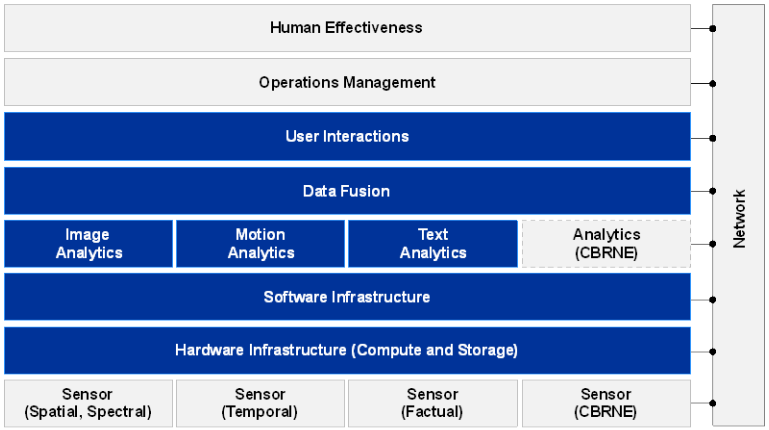
Detect, Track, and Infer Intent of Objects in an Urban Environment with All Source Data



Canonical Decision Support Architecture



Architectural Layers



Operational Issues

- **Shear number of detections, tracklets, track associations overwhelm the limited number of analysts**
- **Integration of tracks from disparate modalities is manually intensive and time consuming**
- **Developing long duration tracks to support social network analysis including patterns of life**
- **Representation of unstructured data that is incomplete, imprecise, uncertain, and contradictory to support analysis, storage and retrieval**
- **Understanding the observed data in the context of multiple hypotheses that are consistent to develop indications and warnings, reduce the number of hypotheses, and to develop new hypotheses**

Meeting mission timelines and operating with large sources of unstructured data requires that the analyst in the loop is more effective

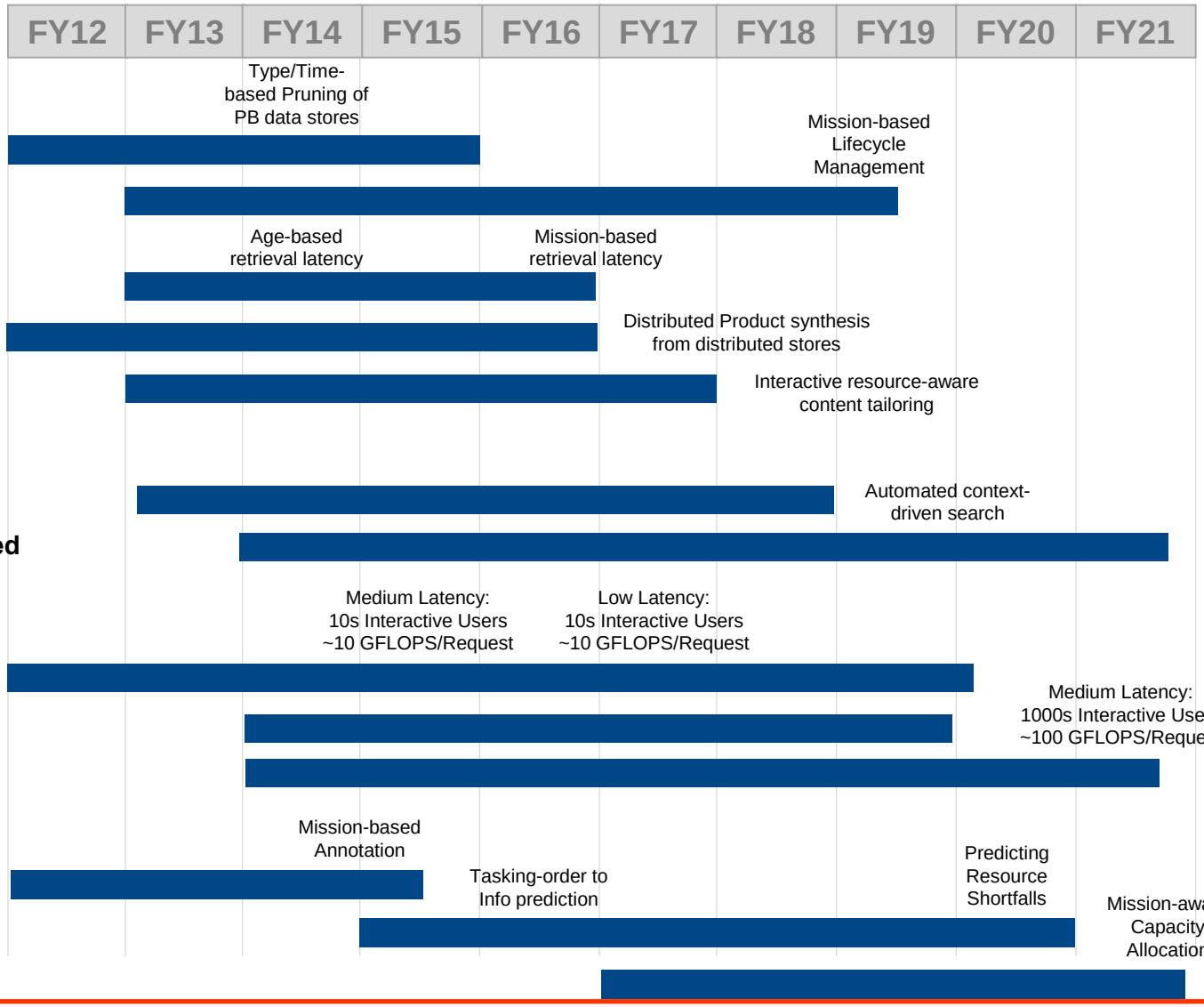
Highest Payoff Capabilities and Associated Metrics

- **Data Management**
  - **Representations: Efficient representation of structured and unstructured data supporting format normalization, mission-aware computation and 100 x compression without loss of fidelity in applications**
- **MOVINT Analysis**
  - **Automated tools that support 100x improvement in the number of tracks that an analyst manages**
    - Probability of correct association of tracklets and tracks > 0.98
    - Time to achieve track association by automation less than current SOA
- **IMINT Analysis**
  - **Automated tools that support 100x improvement in the number of objects, activities, and events that an analyst can manage**
    - Probability of correct classification of objects, activities, and events > 0.98
    - Time to develop objects, activities, and events less than current SOA
- **Text Analysis**
  - **Automated tools that improve by 100 the rate at which information is extracted from documents in any language with**
    - High Probability of correct extraction
- **User Interface**
  - **Automated tools align the information models of all participants in the distributed man machine enterprise that are 98% accurate**

Data Management Layer

- **Problem Statement: Increasing data volumes and modalities have diminished our ability to communicate, store, retrieve and process sources within mission-critical timelines**
- **3-to-5 year timeframe objective**
  - Computational infrastructure to support capturing, processing, marking, retrieval, and management of millions of information objects per second over discovery mission data requirements (PB/TB, long latency)
  - Network architecture with embedded information management on existing networks to support both real-time (MB/GB, low latency) and assisted (GB/TB, medium latency) mission data requirements
- **7-to-10 year timeframe objective**
  - Mission-aware information lifecycle management to age data from (typically) short-term concrete data storage to longer-term symbolic associative representation and retrieval based upon perceived utility and cost
  - Self-balancing merged storage and processing architecture to support analytics with minimal data movement
  - Synchronized anticipatory sensor control and compute/storage resource allocation to support rapid ingest and real-time exploitation

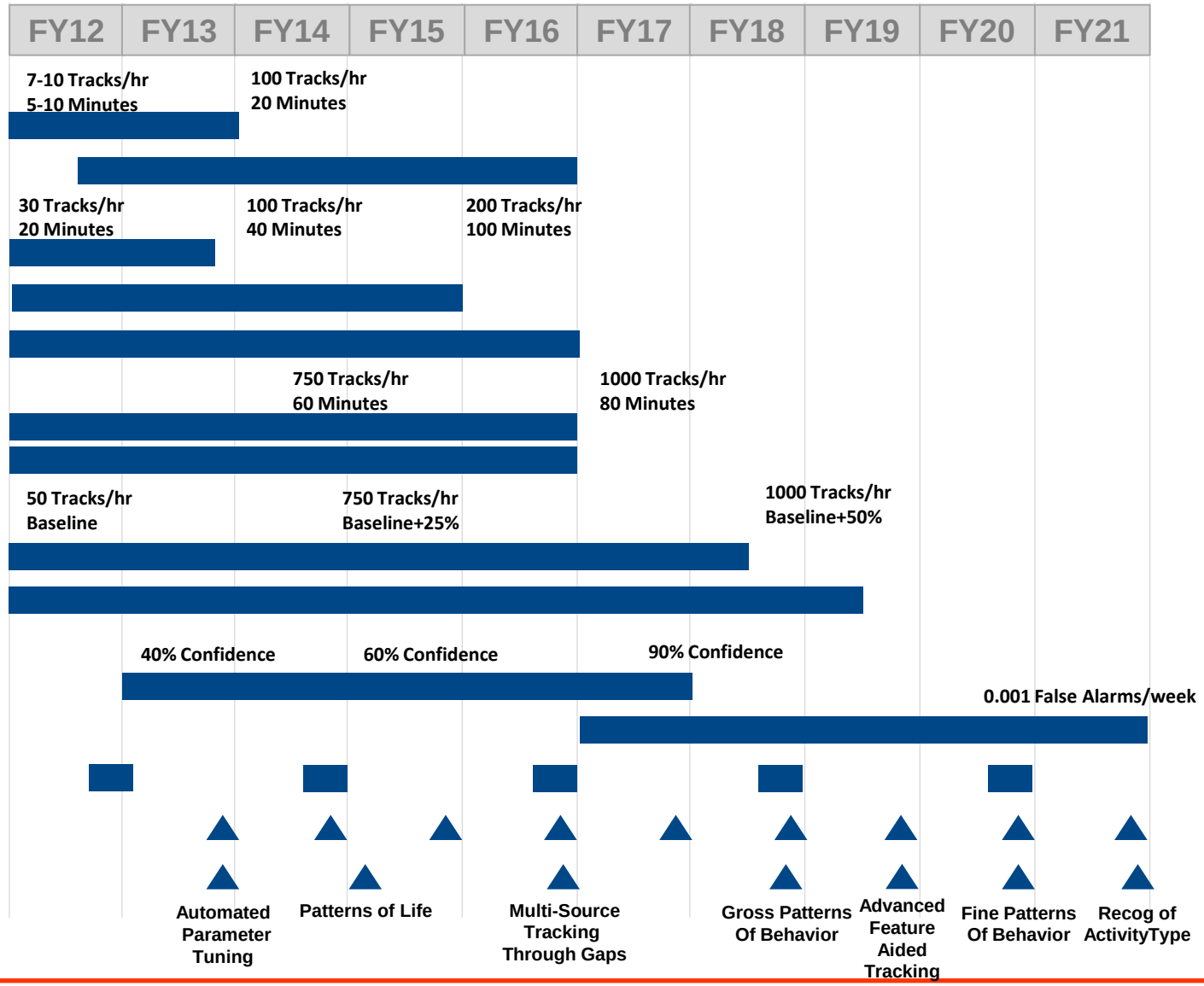
Data Management Roadmap



Analytic Layer

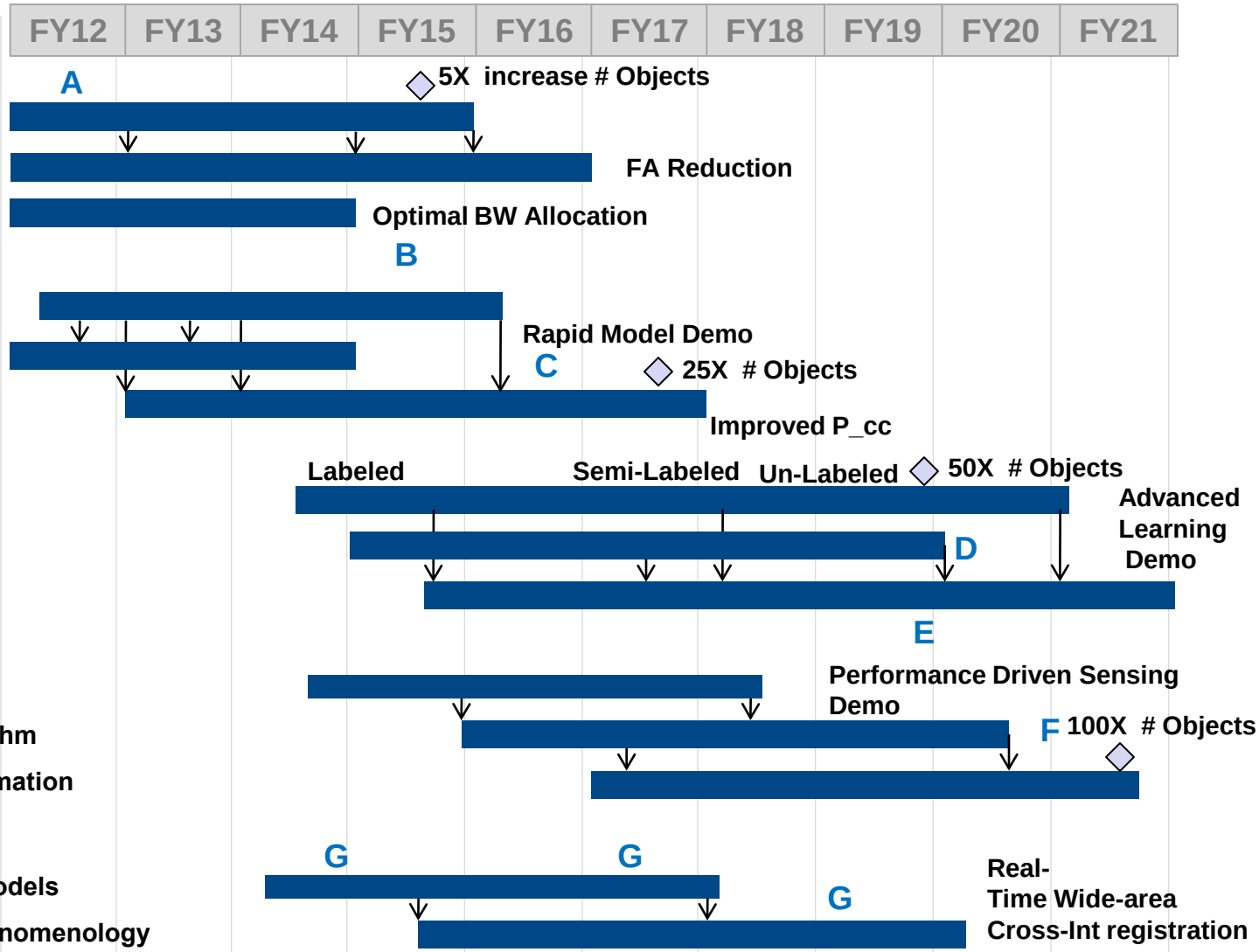
- **Problem Statement: Existing automation tools do not aid users in finding today's complex and adaptable threats within mission timelines**
- **3-to-5 year timeframe objective**
  - Robust classification to accurately detect, geo-register, classify, and identify surface objects despite difficult environments, configurations and emplacements
  - Robust automation tools to identify relationships, patterns of life and activities of objects on the ground
  - Robust tools to capture, store and retrieve HUMINT-based information to identify and leverage popular support against insurgents
  - Domain-specific tools to capture, search, mine and exploit explicit information on insurgent networks from unstructured textual data sources
- **7-to-10 year timeframe objective**
  - Robust automation tools to identify relationships, patterns of life and activities of dismounts
  - Robust tools to search, mine and exploit open-source data to identify all aspects of insurgent networks

MOVINT Roadmap

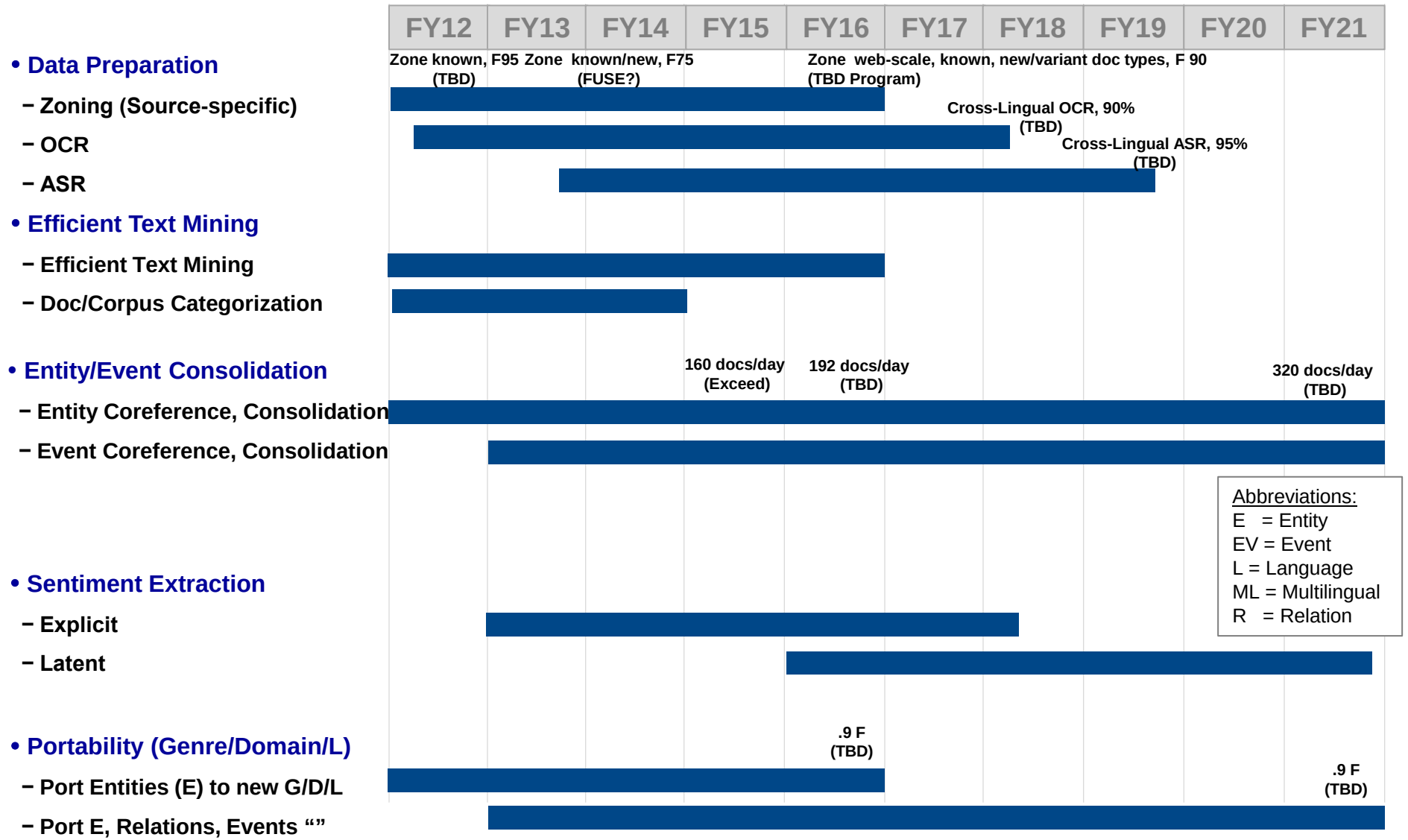


IMINT Roadmap

- **Multi-Source Detection**
  - Precision geo-registration
  - Multi-INT change detection
  - Scalable compression
- **Geometric Features**
  - 3D reconstruction
  - Rapid target insertion
  - Geometric clustering
- **Advanced Learning**
  - Large corpus training
  - Model-based learning
  - On-the-fly adaptation
- **Performance Models**
  - Sensor/Algorithm trade-off
  - Confidence reporting algorithm
  - Predictive performance estimation
- **Accurate Geo-location**
  - Dynamic adaptive sensor models
  - Disparate geometry and phenomenology



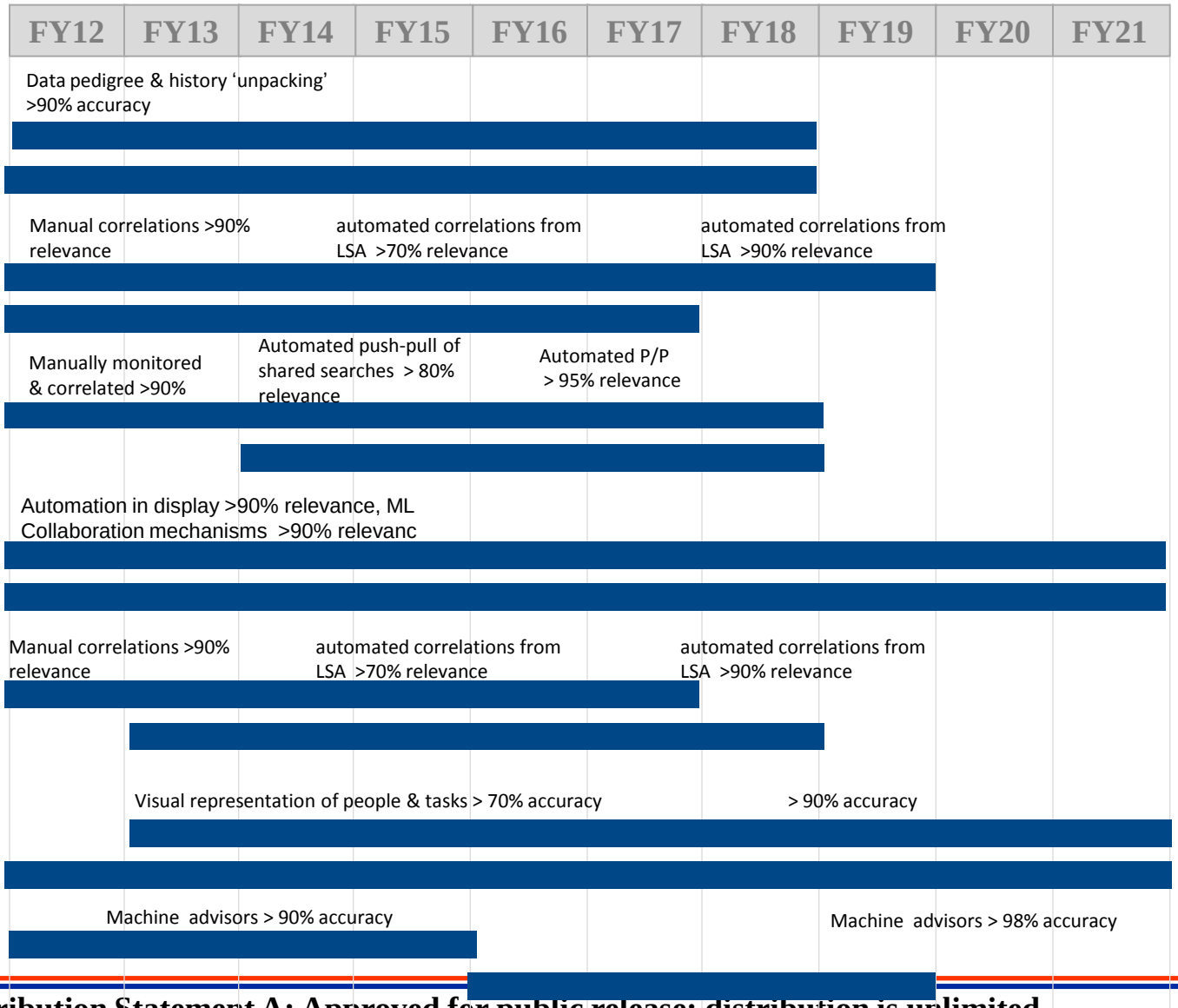
Textual Data Roadmap



User Interaction Layer

- **Problem Statement: Existing interface tools do not detect and proactively respond to the users information needs, given massive amounts of data collected from sensor and open-source assets.**
- **3-to-5 year timeframe objective**
  - Reactive intelligent interfaces
    - Acquisition of massive data, including continuous learning and inference
    - Automatic identification of potential (human) collaborators
    - User-specified interface reconfiguration
  - Adaptable displays that automatically draw human attention to problem areas
  - Workflow tools that guide analysts in complex problems
- **7-to-10 year timeframe objective**
  - Proactive intelligent interfaces and inference engines that:
    - Generate and update rich models of their users current tasks, beliefs and intentions.
    - Socially-guided machine learning to support level 2+ fusion
    - Proactively identify task-relevant data based on current estimates of users beliefs, and intentions and to offer suggestions based on these estimates.
    - Communicate with users in the most natural way possible (language, when appropriate)
  - Workflow tools that capture and teach analysts' best practices

User Interaction Roadmap



FY12 Planned BAA's

- **Data to Decision Special Notice, Spring 2012**
  - POC: Dr. Carey Schwartz
- **ONR Long Range BAA BAA 12-01**
  - POC: Dr. Wen Masters
- **Research Interests of AFOSR BAA 2010-1**
  - POC: Dr. Hugh De Long
- **ARO Core BAA, W911NF-07-R-003-04**
  - POC: Dr. John Lavery
- **DARPA I2O Office Wide BAA, 11-34**
  - POC: Mr. Daniel Kaufman
- **ONR Computational Intelligence for Rapid Accurate Decision Making Special Notice, Spring 2012**
  - POC: Dr. Carey Schwartz

Summary

- **Data representative of the problem domain with ground truth to be made available for development and testing of algorithms**
- **Specifications of a Service Oriented Architecture will be made available to abet government testing and evaluation**
- **Understand the relationship between the "picture" and decisions based upon the picture**
  - Bottoms Up to identify performance controlling functions/modules
  - Top Down to manage quality of picture and manage resources
- **Symbiotic Relationship between automation and humans**
  - Human is cognitive within the architecture and not a servant to the architecture
  - Human mentors the architecture to improve performance
- **Reduce timelines between receipt of data, what does it mean, and what should be done across decision support systems**