Ethical Control of Unmanned Systems

Repeatable Mission Logic and Scenario Simulation will lead to Future Analytic Framework

Don Brutzman
Naval Postgraduate School (NPS)

88th MORS Symposium, M+S Community of Practice

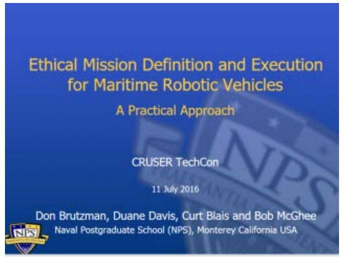
17 June 2020

https://savage.nps.edu/EthicalControl

Topic Abstract … and analytic underpinnings

- **Ethical control of unmanned systems can be accomplished through structured mission definitions that are consistently readable, validatable and understandable by humans and robots.**
  - Structured data models for syntactically valid, well-defined mission orders.
  - Semantic Web representations and queries to ascertain semantic correctness.
- **Modeling of scenarios and simulations that test ethical constraints can lead to visualization of outcomes.**
  - Readable by humans, conduct probabilistic analysis, Web-based visualization.
  - Repeatable in any programming language or robot dialect for scalability.
- **We foresee analytic foresight and assessment leading to human-driven qualification of systems possessing lethal/lifesaving force.**
  - All about the data! Can we read, repeat, replay, review, remix, understand?

Key Insights regarding Human Ethical Control



1. **Humans in military units are able to deal with moral challenges without ethical quandaries,**
   - **by using formally qualified experience, and by following mission orders that comply with Rules of Engagement (ROE) and Laws of Armed Conflict (LOAC)**.

2. Ethical behaviors don’t define the mission plan. Instead, **ethical constraints inform the mission plan**.

3. **Naval forces can only command mission orders that are**
   - **Understandable by (legally culpable) humans, then**
   - **Reliably and safely executed by robots**.

Reference: CRUSER TechCon Overview 2016
https://gitlab.nps.edu/Savage/EthicalControl/tree/master/documents/presentations

Example AVCL mission agenda, as pseudo-code XML

```xml
<?xml version="1.0" encoding="UTF-8"?>
<UUVMission>
   <GoalSet>
      <Goal area=”A” id=”goal1”>
         <Search nextOnSuccess=”goal2” nextOnFailure=”goal3”/>
      </Goal>
      <Goal area=”A” id=”goal2”>
         <SampleEnvironment nextOnSuccess=”goal3”
                nextOnFailure=”recover”/>
      </Goal>
      <Goal area=”B” id=”goal3”>
         <Search nextOnSuccess=”goal4” nextOnFailure=”goal4”/>
      </Goal>
      <Goal area=”C” id=”goal4”>
         <Rendezvous nextOnSuccess=”recover” nextOnFailure=”recover”/>
      </Goal>
      <Goal area=”recoveryPosition” id=”recover”>
         <Transit nextOnSuccess=”missionComplete”
                nextOnFailure=”missionAbort”/>
      </Goal>
   </GoalSet>
</UUVMission>
```

AVCL is readable by human or robot, captures logic of mission tasking

XML ensures syntactically correct, well-defined, numerically valid

Needed: semantic representation to check ethical, logical consistency

AVCL mission goals vocabulary (Davis 2015)

AVCL mission goals	Define	Used	Definition
Attack	partial	√	To conduct a type of offensive action characterized by employment of firepower and maneuver to close with and destroy an enemy.
Decontaminate	√		To provide purification making an area safe by absorbing, destroying, neutralizing, making harmless, or removing chemical, biological, or nuclear contamination.
Demolish	√		To destroy structures, facilities, or material by any available means.
IlluminateArea	√		To provide locale lighting by searchlight or pyrotechnics.
Jam	√		To deliberately radiate, re-radiate or reflect electromagnetic energy with the object of impairing the use of electronic devices or systems.
MarkTarget	√	√	To make visible (by the use of light, infrared, laser, smoke, etc.) of an object in order to allow its identification by another object.
MonitorTransmissions	√	√	To conduct electronic warfare support operations with a view to searching, locating, recording and analyzing radiated electromagnetic energy.
Patrol	√	√	To gather information or carry out a security mission.
Rendezvous	√	Partial	Achieve a meeting at a specified time and place.
Reposition	√	√	To change position from one location to another.
SampleEnvironment	Partial	√	Collect environmental samples for testing for chemical compounds, biological creatures, or nuclear hazards.
Search	√	√	To look for lost or unlocated objects or persons.

More Goal Types Foreseen

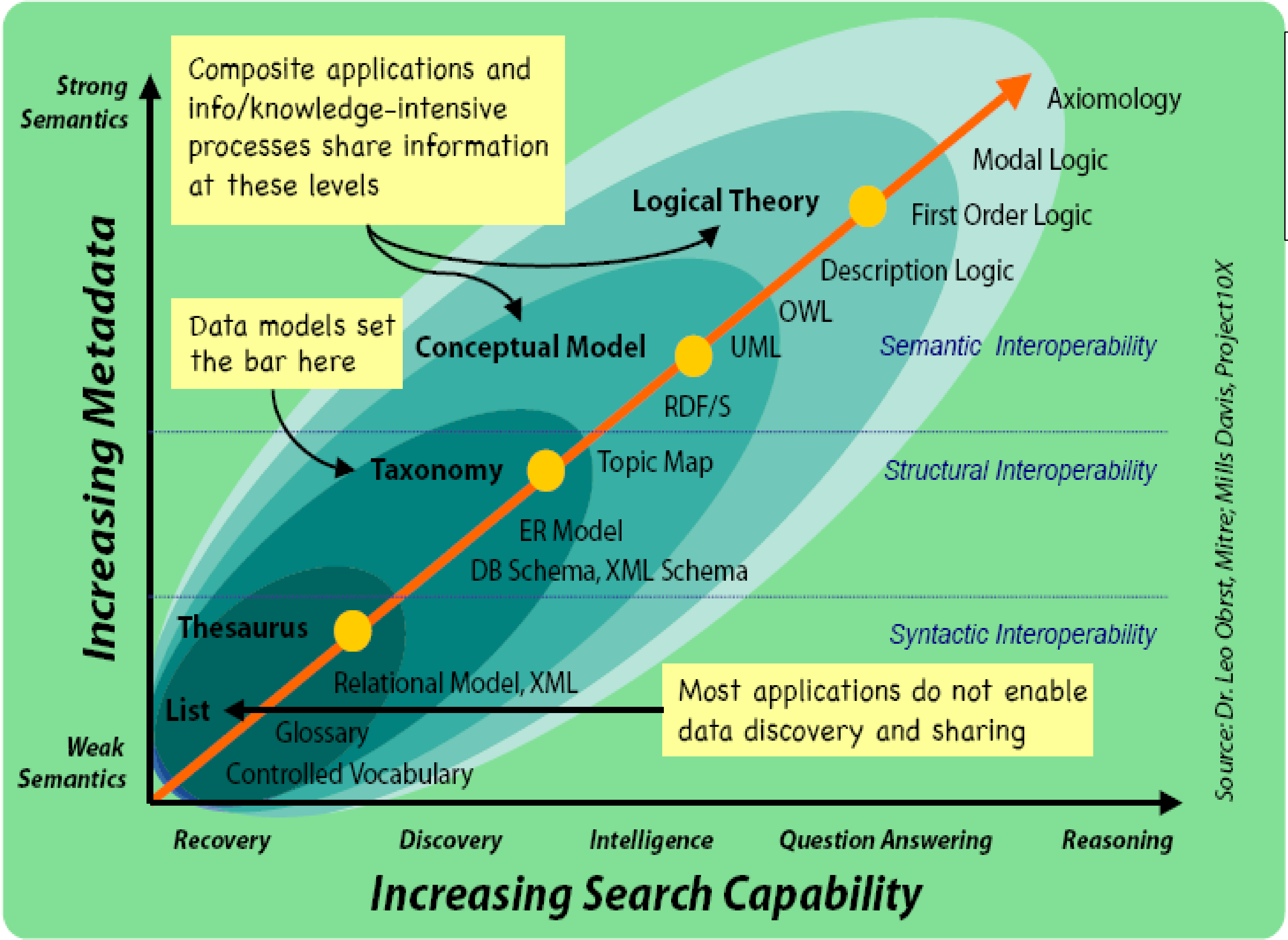
Synopsis: Ethical Control of Unmanned Systems

- **Project Motivation**: ethically constrained control of unmanned systems and robot missions by human supervisors and warfighters.
- **Precept:** well-structured mission orders can be syntactically and semantically validated to give human commanders confidence that offboard systems
  - *will **do** what they are told to **do***, and further
  - *will **not do** what they are **forbidden to do***.

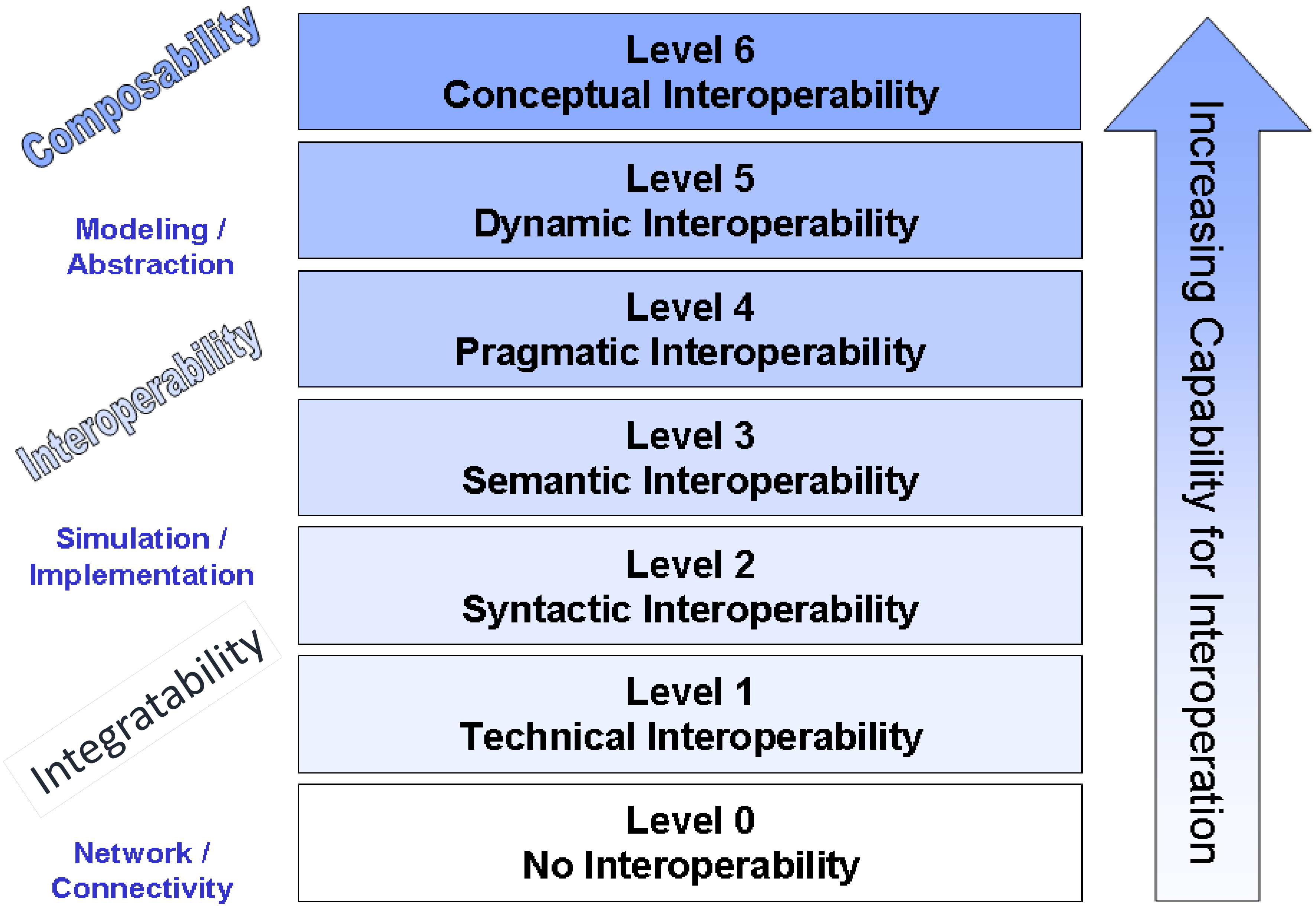
Paraphrase: can qualified robots correctly follow human orders?

- **Project Goal:** apply Semantic Web ontology to scenario goals and constraints for logical validation that human-approved mission orders for robots are semantically coherent, precise, unambiguous, and without internal contradictions.
- **Long-term Objective:** demonstrate that no technological limitations exist that prevent applying the same kind of ethical constraints on robots and unmanned vehicles that already apply to human beings.

Improving Semantic Representation



Improving Interoperability: Levels of Conceptual Interoperability Model



Semantic Web Stack

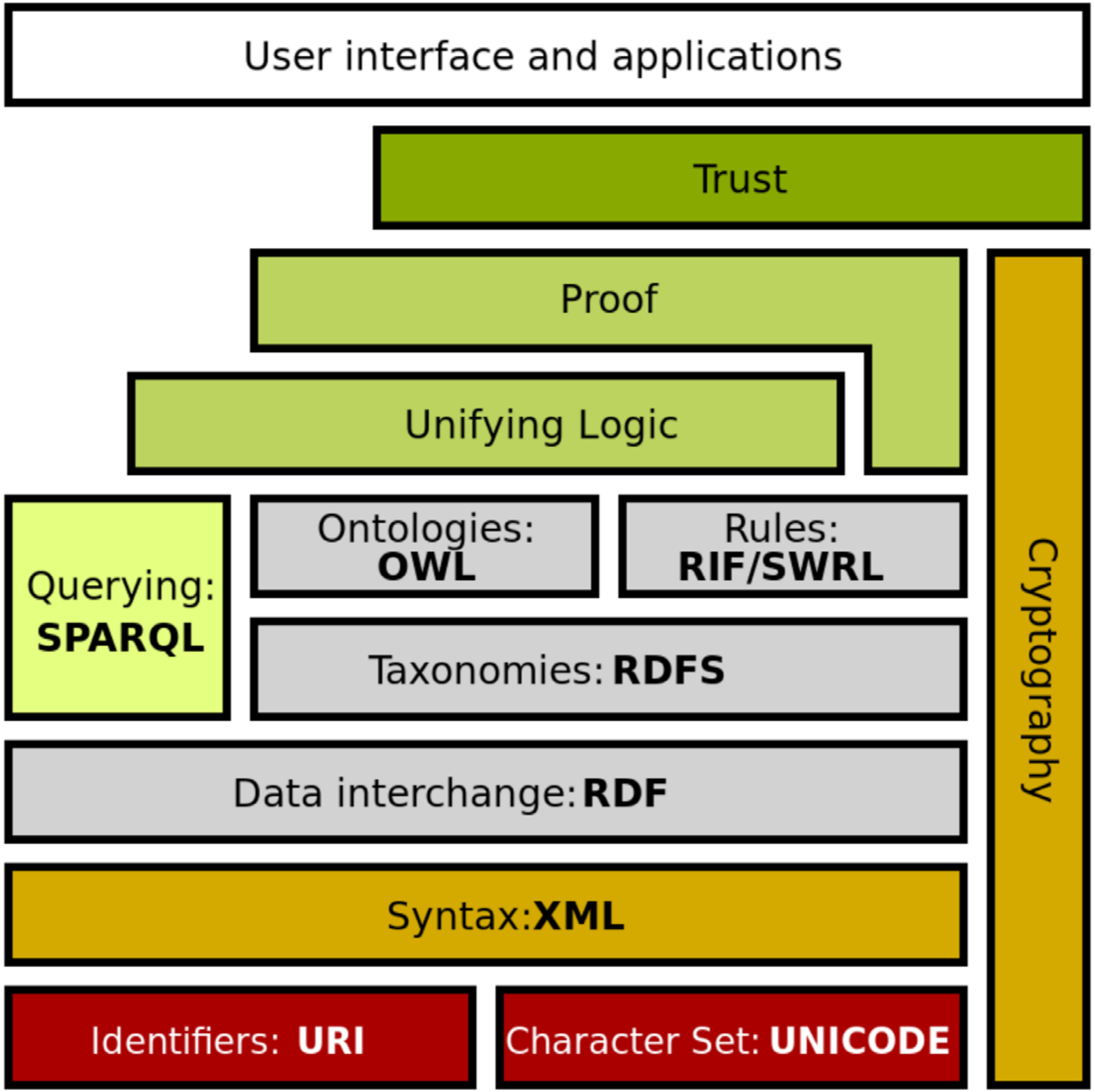
Extends larger Web architecture

- All of these data languages are approved W3C standards
- Proof and unifying logic are mathematically well defined

Trusting derived (composed) statements arises from

- Encryption + digital signature confirms trusted data sources
- Formal logic is basis for deriving new information
- Wikipedia: [Semantic Web Stack]

Of note: this project is exercising every layer of Semantic Web stack.



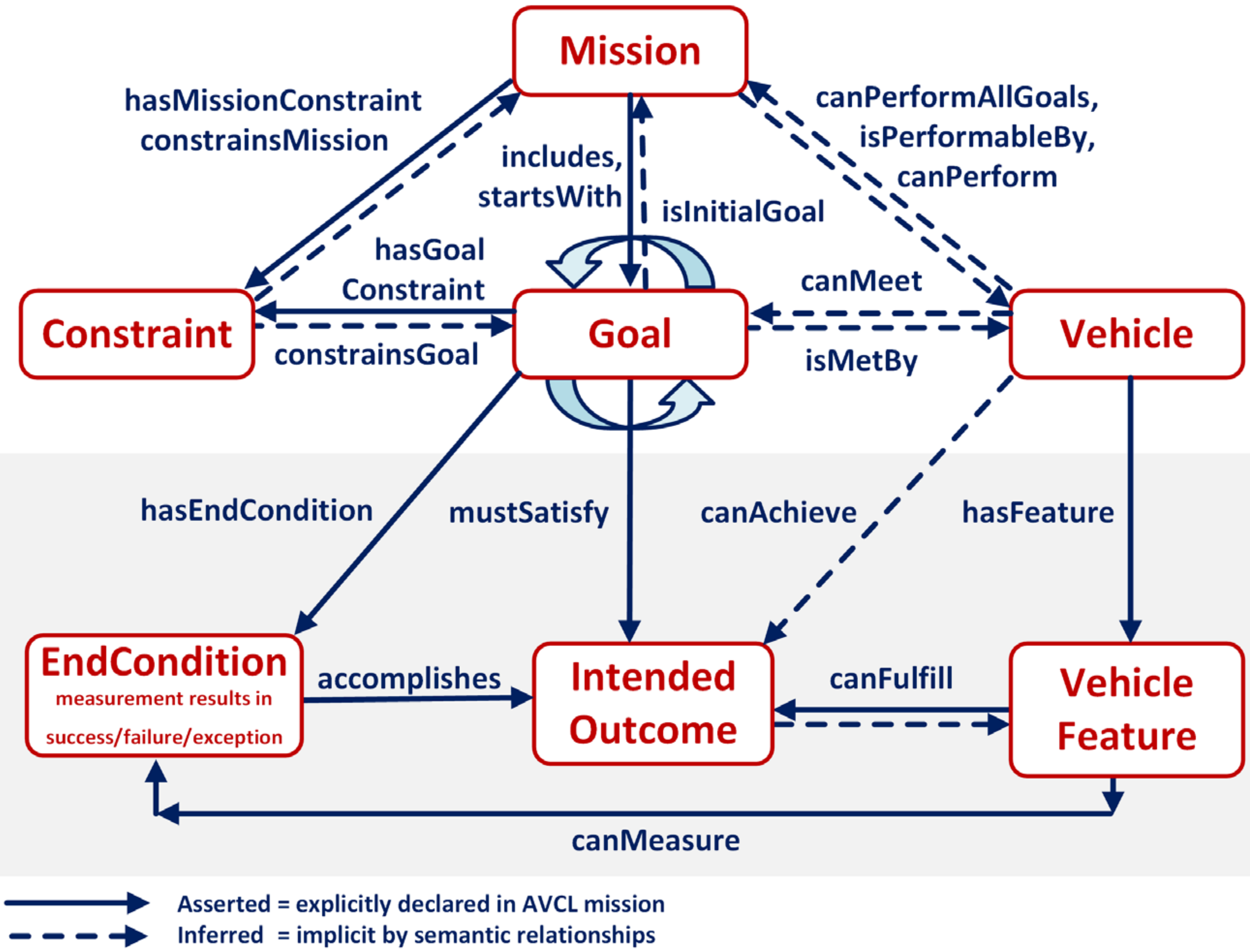
Mission Execution Ontology (MEO) for Ethical Control of Unmanned Systems in Surrogate Scenarios

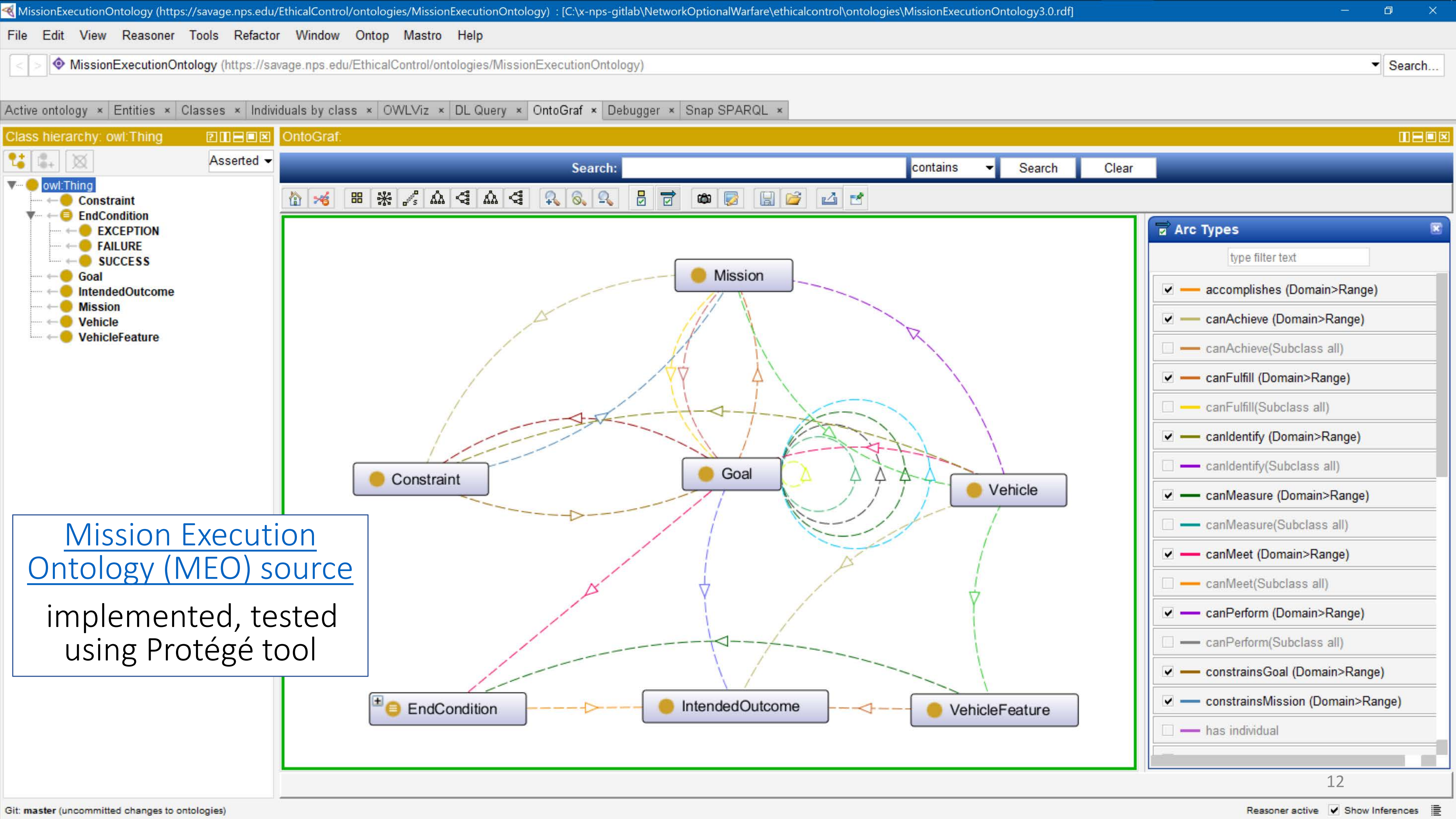
Summary of relationships

- **Autonomous Vehicle Command Language (AVCL) for Missions.**
  - Declarative XML, years of NPS research.
- **Multiple Mission Representations.**
  - Imperative commands (orders/waypoints/etc.).
  - Declarative commands (mission goals).
  - Mission results (order log, telemetry etc.).
  - Mission metadata for parameters, settings.
  - Lisp and Prolog examples (Bob McGhee, NPS).
- **Autonomous Unmanned Vehicle (AUV) Workbench Simulation and Visualization Support**
  - Recently restored, debug testing commenced.
  - AVCL 2.1 is prior published version, centered on *syntactic validation*, solo robot operations.
  - AVCL 3.0 is new working version for testing range of multi-participant missions.

- **Mission Execution Ontology (MEO) for Semantic Validation**
  - Semantic Web framework of rules, relationships for *ethical validation.*
  - Initial examples in IEEE JOE paper.
  - Retested using current Protégé, Jena tools.
- **Sailor Overboard and Other Missions**
  - Hand-crafted triples using Turtle syntax.
  - Beginning to build unit testing framework.
  - Confirming correlation of AVCL information model to existing MEO ontology.
  - Automatic conversion of AVCL missions to match, thus accelerating multiple-mission testing on diverse systems.
  - Visualization, reporting via AUV Workbench can aid understanding, mission planning and further progress.

Mission Execution Ontology (MEO) 3.0

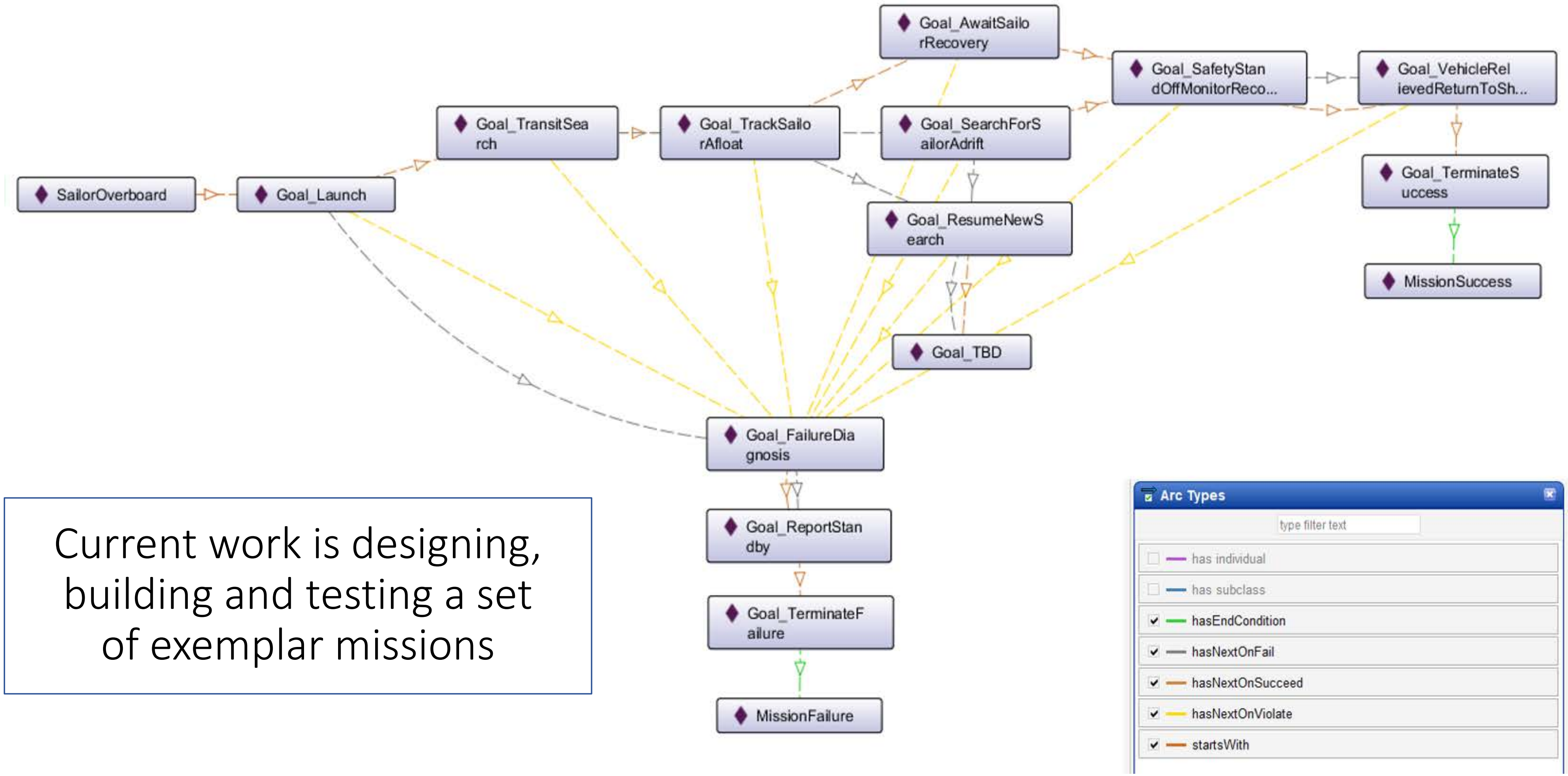




[Mission Execution Ontology (MEO) source]

implemented, tested using Protégé tool

Ethical Control of Unmanned Systems in a Surrogate Scenario: Sailor Overboard Mission defined using the MEO Ontology



Current work is designing, building and testing a set of exemplar missions

Core Considerations for Artificial Intelligence (AI)

- Effective AI turns data into information for use by humans.
- AI systems do not have capacity for rational thought or morality.
- Unmanned systems require sophisticated control across time, space.
- A large and involved body of internationally accepted law comprises Law of Armed Conflict (LOAC), bounding Rules of Engagement (ROE).
- Only professional warfighters have moral capacity, legal culpability, and societal authority to direct actions applying lethal force.
- Humans must be able to trust that systems under their direction will <u>*do what they are told to do*</u>, and <u>*not do what they are forbidden to do*</u>.
- Successful Ethical Control of unmanned systems must be testable.

Mission clarity for humans – and robots

- Simplicity of success, failure, and (rare) exception outcomes encourages well-defined tasks and unambiguous, measurable criteria for continuation.

Confirmable beforehand: can a tactical officer (or commanding officer) review such a mission and then confidently say

  - “yes I understand and approve this human-robot mission” or, equivalently,
  - “yes I understand this mission and my team can carry it out themselves.”

Converse:

- if an officer can’t fully review/understand/approve such a mission, then likely it is **ill-defined** and needs further clarification anyway.

Added benefit: missions that are clearly readable/runnable by humans and robots can be further composed and checked by C2 planning tools to test for group operational-space management, avoiding mutual interference, etc.

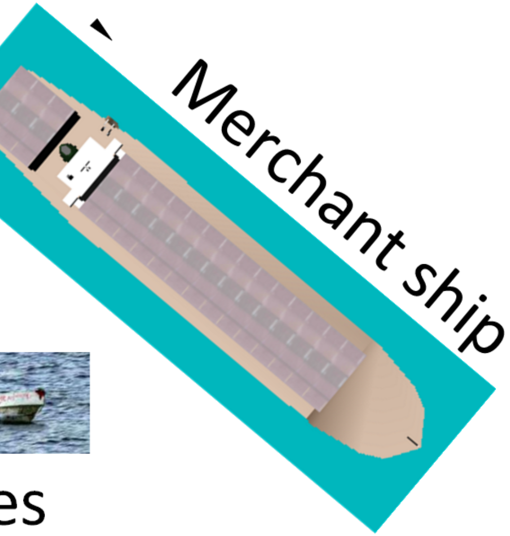
Life boat

Life-saving force: locate, track, communicate, beacon

Ethical control of unmanned systems is required for both lethal and lifesaving force if remote robots communicate intermittently, operating across lengthy time and distance.

Response dilemma for U.S. Navy ship

Lethal force: locate, warn, defend, threaten, attack



Pirates

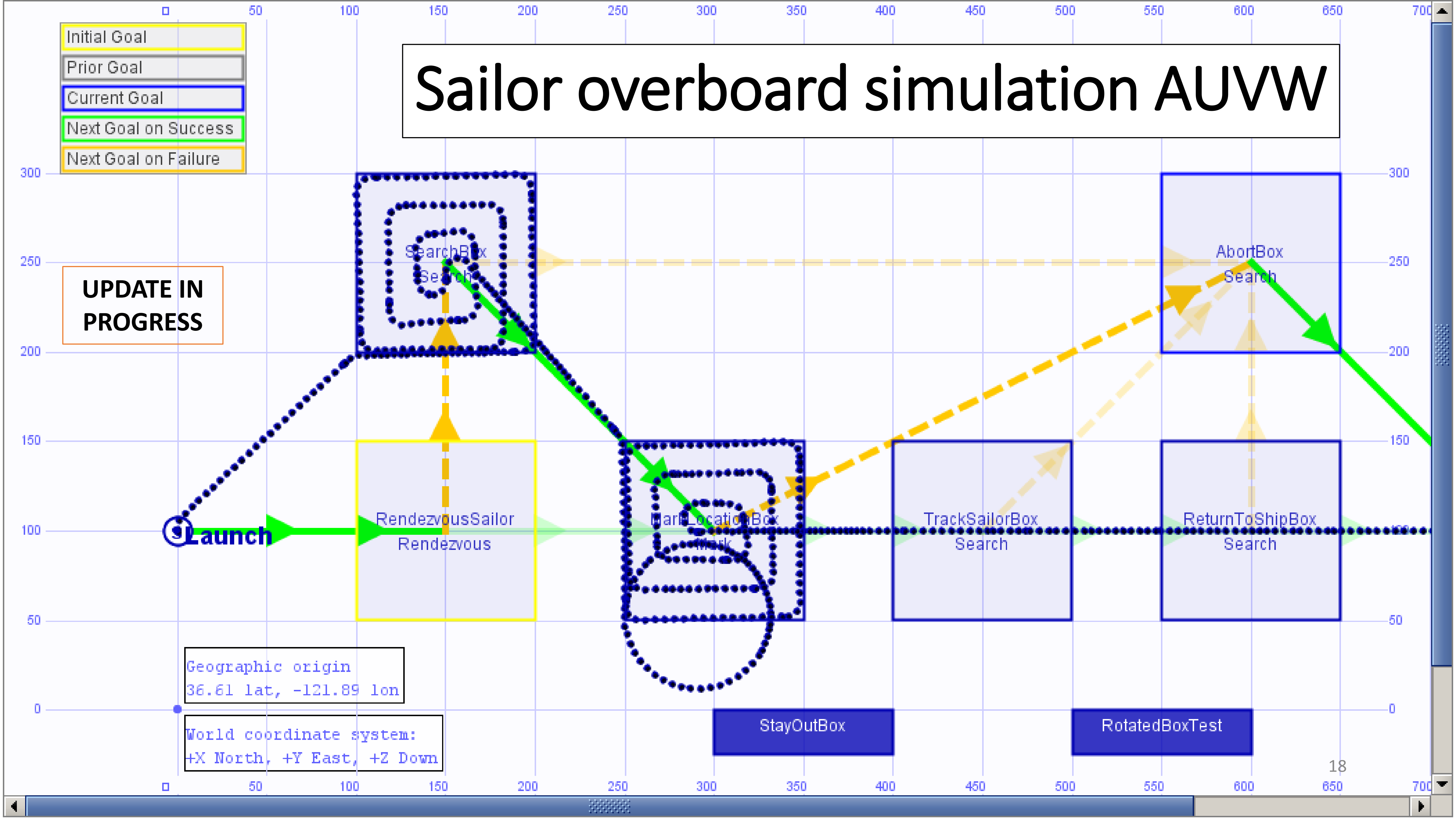
Respond to one or both scenarios with USV/UAV assets to establish on-scene visibility and presence

OODA Loops for Ethical Control Canonical Missions

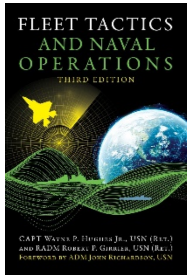
Ethical Control OODA Loops	Observe	Orient	Decide	Act
Sailor Overboard	Find Sailor	Report status	Avoid interference	Track sailor until rescued or relieved
Lifeboat Rescue	Find Lifeboat	Report status	Two-way communication	Track life raft until relieved
Pirate Seizure of Merchant Ship	Find merchant ship, pirate small boats	Identity Friend Foe Neutral Unknown (IFFNU) Issue warnings	Human commander authorization to use lethal force	Attack to defend ship if provoked, stay with merchant
Hospital Ship Swarm Attack	EM threat signals detected	(no orientation step in Sense Decide Act)	Reflex-response weapons attack	Mistaken attack on friendly = war crime
Hospital Ship Defense detects spoofing anti-pattern	EM threat signals detected	IFFNU including correlation	Human requirement for lethal force unmet, attack avoided	Report threat alert, commence search for hostile actors

Sailor overboard simulation AUVW

UPDATE IN PROGRESS



Fleet Tactics and Naval Operations



Wayne P. Hughes, Jr. and Robert P. Girrier, *Fleet Tactics and Naval Operations*, Third Edition, Naval Institute Press, Annapolis Maryland, June 2018.

- https://www.usni.org/press/books/fleet-tactics-and-naval-operations-third-edition

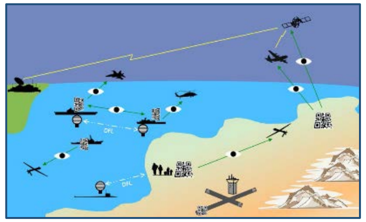
From newly added Chapter 12, A Twenty-First-Century Revolution:

- "At the most fundamental level, [Information Warfare] IW is about how to employ and protect the ability to sense, assimilate, decide, communicate, and act – while confounding those same processes that support the adversary."
- "Information Warfare broadly conceived is orthogonal to naval tactics. As a consequence, IW is having major effects on all six processes of naval tactics used in fleet combat – scouting and antiscouting, command-and-control, C2 countermeasures, delivery of fire, and confounding enemy fire."
- "Indeed there is a mounting wave of concern about how far automation will expand and what its impact will be on the continuum of cognition from data to information to knowledge. […] Navies are facing similar uncertainties."

Wayne Hughes coined the term "Network Optional Warfare" after many discussion sessions, directly contrasting it to Network Centric Warfare. Thank you sir.

Network Optional Warfare (NOW)

Overall Concept



Naval forces do not have to be engaged in constant centralized communication. Deployed Navy vessels have demonstrated independence of action in stealthy coordinated operations for hundreds of years.

- Littoral operations, deployable unmanned systems, and a refactored force mix for surface ships pose a growing set of naval challenges and opportunities. Network-optional warfare (NOW) precepts include Efficient Messaging, Optical Signaling, Semantic Coherence and Ethical Human Supervision of Autonomy for deliberate, stealthy, minimalist tactical communications.
- https://wiki.nps.edu/display/NOW/Network+Optional+Warfare

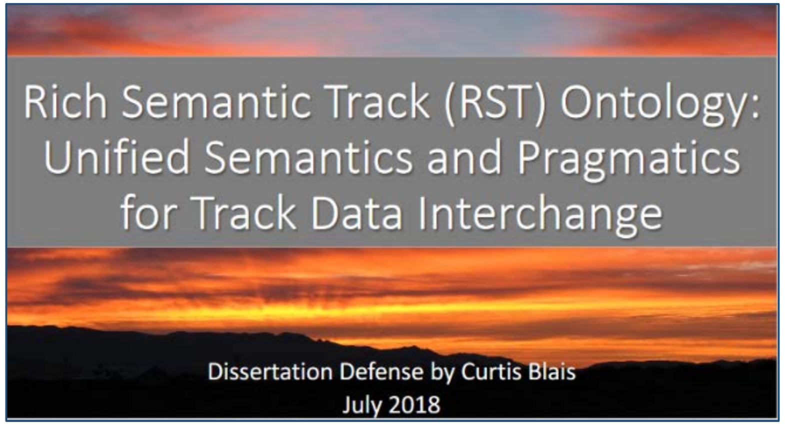
Rich Semantic Track (RST)

Related Work

Curt Blais Dissertation

- DoD mandates data sharing practices, but practices have been mixed and uneven, resulting in perpetuation of system-centric data practices.
- Sharing and collective understanding of track data - collections of time-stamped perceptions of the state of objects of interest — are critical to warfighting systems.

- Shared understanding requires common semantics.
- The Rich Semantic Track (RST) ontology provides a foundation for shared understanding of track data.
- It is time to change the way DoD manages data and engineers systems, starting with adoption of the RST ontology and moving toward the vision of a Web of linked track data.



[Network Optional Warfare (NOW) Semantic Coherence]

[Rich Semantic Track (RST)]

Assessment of Current Thinking in Ethical AI

- Human supervision of potentially lethal autonomous systems is a matter of serious global importance.
- Wide consensus is emerging on principles, aspects of the problem, elements of solutions, and need to achieve better capabilities.
- Much philosophical concern but few concrete activities are evident.

Ethical Control of Unmanned Systems project appears to provide a needed path towards practice, with the historic role of warfighting professionals more central than ever as weapons autonomy grows.

Conclusions

- Human supervision is required for any unmanned systems holding potential for lethal force.
  - Cannot push "big red shiny AI button" and hope for best – immoral, unlawful.
  - Similar imperatives exist for supervising systems holding life-saving potential.
- Human control of unmanned systems is possible at **long ranges of time-duration and distance** through well-defined mission orders.
  - Readable and sharable by both humans and unmanned systems.
  - Validatable syntax and semantics through understandable logical constraints.
  - Testable and confirmable using simulation, visualization, perhaps qualification.
- Coherent human-system team approach is feasible and repeatable.
  - Semantic Web confirmation can ensure orders are comprehensive, consistent.
  - Human role remains essential for life-saving and potentially lethal scenarios.

Recommendations for Future Work

Continued development

- Diverse mission exemplars
- Software implementations
- 2D, 3D visualization of results

Future capabilities

- Automatable testing
- Field experimentation (FX)
- “Qualification” of unmanned systems in virtual environments

Outreach

- Presentations, publication review
- Engagement in key ethical forums
- NPS wargame and course support
- NPS thesis and dissertation work

Adoption

- Support for developmental system
- Influence campaign for both C4I and robotics communities of interest

Analytic questions of interest – to warfighters

Repeatability

- What happened? What changed? What's next? What will happen?
- What can we learn from all data + all operations? … Will we prevail?

Understandability

- Do I entrust this robot with authority for lethal or lifesaving force?
- Do these combined human-machine actions make sense to team?

Trust and Verification, Validation, Accreditation (V V+A)

- Is a robot system (hardware, software, sensors) qualified to deploy?
- What is Modeling + Simulation basis for analysis-development loop?

Contact

Don Brutzman

brutzman@nps.edu
http://faculty.nps.edu/brutzman

Code USW/Br, Naval Postgraduate School

Monterey California 93943-5000 USA

1.831.656.2149 work

1.831.402.4809 cell

CRUSER · NEWS

Consortium for Robotics and Unmanned Systems Education and Research

ETHICAL MISSION DEFINITION AND EXECUTION FOR MARITIME ROBOTS UNDER HUMAN SUPERVISION

- Lethality requires ethical and legal basis, supervised by military teams.
- Executable robot tasking can resemble tactical tasking of humans afloat.
- Careful application of goal constraints makes ethical control feasible.
- Robot missions then complement and extend naval operation orders.
- Semantic Web logic can confirm ethical correctness and completeness.
- Next steps: continue 2 decades of work with realistic scenario testing.

> "Ethical constraints on robot mission execution are possible today. There is no need to wait for future developments in Artificial Intelligence (AI). It is a moral imperative that ethical constraints in some form be introduced immediately into the software of all robots that are capable of inflicting unintended or deliberate harm to humans or property." Robert McGhee, April 2016

- IEEE Journal of Oceanic Engineering (JOE) paper along with online references.
- Authors Don Brutzman, Curtis Blais, Duane Davis and Robert McGhee, NPS.
- Feedback and recommendations always welcome. Contact: brutzman@nps.edu