Ethical Control of Unmanned Systems

Mission Design and Semantic Web Exemplars for Human Supervision of Lethal/Lifesaving Autonomy

Don Brutzman and Curt Blais, Naval Postgraduate School (NPS)

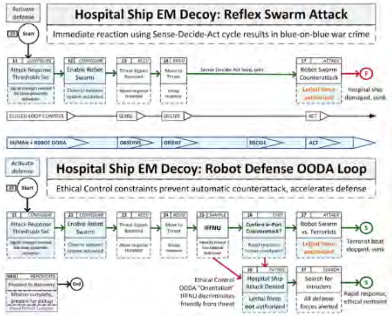
Unmanned Vehicle and Autonomous Systems (UVAS)

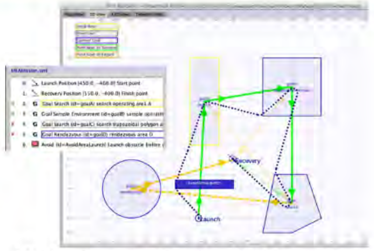
Naval Working Group Meeting

25 March 2020

Ethical Control of Unmanned Systems:
Keeping Warfighters in Charge of Autonomy

CRUSER
Consortium for Robotics and Unmanned Systems Education and Research





Mission Execution Ontology (MEO) evolving

Milestones and Transitions

- CRUSER development led to first project selection under CRADA with Raytheon Missile Systems (RMS).
- Successful progress on test missions entering TRL 5 with simulation and Web-sharable 3D visualization.
- Expressing multiple robot mission plans consistently, coherently for diverse UAV, USV, UUV platforms.
- Use Semantic Web Standards to support warfighters.
- Evaluate NAVSEA Unmanned Maritime Autonomy Architecture (UMAA) evolution for robot qualification.

Why / Objectives

- Ethical control of unmanned systems can be accomplished through structured mission definitions that are trusted, consistently readable, validatable, repeatable and understandable by humans and robots.
- Orders must be lawful. Unmanned systems must behave ethically and comprehensibly if they are to support manned military units effectively.
- Well-structured mission orders can be tested and trusted 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*.
- Demonstrate that no technological limitations exist that prevent applying the same kind of ethical constraints on robots and unmanned vehicles that already apply to humans, in lethal and life-saving scenarios.

https://savage.nps.edu/EthicalControl

What / Deliverables

- Update Mission Execution Ontology (MEO) concepts demonstrated in tests and simulation, building to perform field experimentation (FX).
- Supervise thesis work to explore canonical exemplar missions that are expected to utilize unmanned systems, looking across the full range of Naval warfare communities. Example scenarios include UAV for sailor overboard, UAV for refugee/lifeboat escort, and adept scouts. All must observe Law of Armed Conflict (LOAC), Rules of Engagement (ROE), and moral guidance of commanders despite long durations/distances.
- Define, simulate, and test combination of real-world goals and ethical constraints to robot mission tasking across set of canonical scenarios.
- Illustrate how human-robot teams meet moral and legal requirements if deploying unmanned systems with potential for lethal, life-saving force.

Building on Concept Demonstrations

Principal Investigator: Don Brutzman brutzman@nps.edu 1.831.656.2149
Co-Investigator: Curtis Blais clblais@nps.edu 1.831.656.3215

Synopsis: Ethical Control of Unmanned Systems

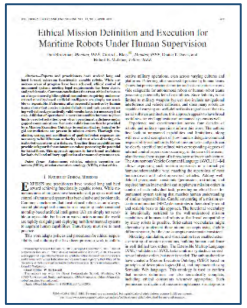
- **Project Motivation**: ethically constrained control of unmanned systems and robot missions by human supervisors and warfighters.
- **Precept:** well-structured mission orders can be syntactically and semantically validated to give human commanders confidence that offboard systems
  - *will **do** 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.

https://savage.nps.edu/EthicalControl

Ethical Mission Definition and Execution for Maritime Robots Under Human Supervision


Ethical Mission Definition and Execution for
Maritime Robots Under Human Supervision

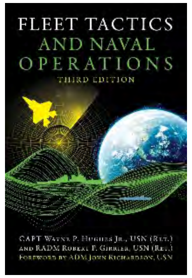
Theoretical Basis

- Don Brutzman, Curtis L. Blais, Duane T. Davis, Robert B. McGhee

- [IEEE Journal of Oceanic Engineering (JOE), Volume: 43 , Issue: 2 , April 2018]

- *Abstract.* Experts and practitioners have worked long and hard toward achieving functionally capable robots. While numerous areas of progress have been achieved, ethical control of unmanned systems meeting legal requirements has been elusive and problematic. Common conclusions that treat ethical robots as an always-amoral philosophical conundrum requiring undemonstrated morality-based artificial intelligence are simply not sensible or repeatable. Patterning after successful practice by human teams shows that precise mission definition and task execution using well-defined, syntactically valid vocabularies is a necessary first step. Addition of operational constraints enables humans to place limits on robot activities, even when operating at a distance under gapped communications. Semantic validation can then be provided by a Mission Execution Ontology to confirm that no logical or legal contradictions are present in mission orders. Thorough simulation, testing, and certification of qualified robot responses are necessary to build human authority and trust when directing ethical robot operations at a distance. Together these capabilities can provide safeguards for autonomous robots possessing the potential for lethal force. This approach appears to have broad usefulness for both civil and military application of unmanned systems at sea.

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.

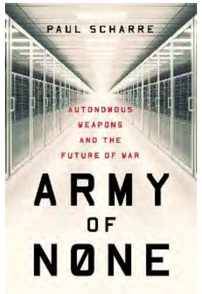
- https://www.usni.org/press/books/fleet-tactics-and-naval-operations-third-edition

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.

Army of None: Autonomous Weapons and the Future of War



Paul Scharre, *Army of None: Autonomous Weapons and the Future of War,* W.W. Norton, New York 2018. www.paulscharre.com/army-of-none

- "What happens when a Predator drone has as much autonomy as a Google car? Or when a weapon that can hunt its own targets is hacked? Although it sounds like science fiction, the technology already exists to create weapons that can attack targets without human input."
- *"Army of None* engages military history, global policy, and cutting-edge science to argue that we must embrace technology where it can make war more precise and humane, but without surrendering human judgment. When the choice is life or death, there is no replacement for the human heart."
- Interestingly anticipates many of the approaches taken in this project.

Autonomous Vehicle Command Language (AVCL)

Mission Tasking

- AVCL is a command and control language for humans supervising autonomous unmanned vehicles.
  - Clarity arises from close correspondence to human naval terminology.
- Structured vocabulary defining terms and relationships for mission planning, execution, conduct, recording and replay across diverse robot types.
- Common-ground XML representations for
  - Mission agenda plans, mission scripts, and post-mission recorded telemetry results.
  - ***Future work:*** defining unit tests and expected results for verification and validation.
- Operators have single archivable, validatable format for robot tasking, results
  - directly convertible to and from a wide variety of different robot command languages.

https://savage.nps.edu/Savage/AuvWorkbench/AVCL/AVCL.html

Example AVCL mission agenda, as pseudo-code 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

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.

Wrong question, right question

Wrong question to ask first when planning a tactical operation:

- “What are my *robots* doing out there?”

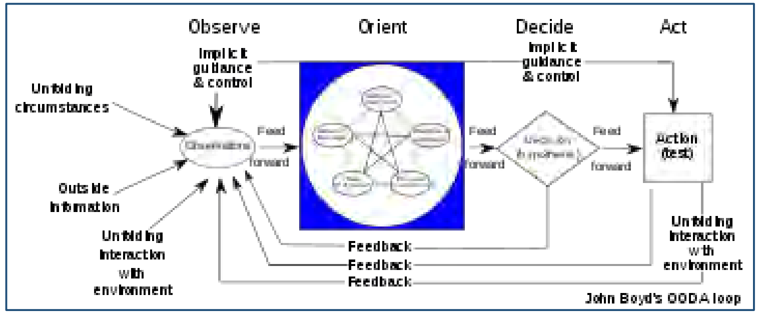
Right question to ask first when planning a tactical operation:

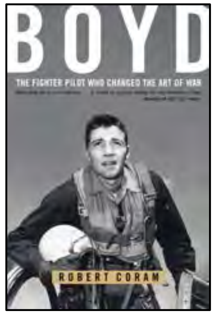
- “What is my *human-robot team* doing out there?”

Human-robot team mission has to be understood first!

- Robots complement humans, who must remain in charge throughout.
- If you don’t have an OODA loop, you don’t have a competent plan.

John Boyd and OODA Loop





Wikipedia: John Boyd (military strategist) and Observe Orient Decide Act (OODA) Loop

- "… the key to victory is to be able to create situations wherein one can make appropriate decisions more quickly than one's opponent."

Robert Coram, *BOYD: The Fighter Pilot Who Changed the Art of War*, 2004.

- "John Boyd may be the most remarkable unsung hero in all of American military history. Some remember him as the greatest U.S. fighter pilot ever - the man who, in simulated air-to-air combat, defeated every challenger in less than forty seconds. Some recall him as the father of our country's most legendary fighter aircraft - the F-15 and F-16. Still others think of Boyd as the most influential military theorist since Sun Tzu. They know only half the story."
- "Boyd, more than any other person, saved fighter aviation from the predations of the Strategic Air Command. His manual of fighter tactics changed the way every air force in the world flies and fights. He discovered a physical theory that forever altered the way fighter planes were designed. Later in life, he developed a theory of military strategy that has been adopted throughout the world and even applied to business models for maximizing efficiency. And in one of the stories of modern military history, the Air Force fighter pilot taught the U.S. Marine Corps how to fight war on the ground. His ideas led to America's swift and decisive victory in the Gulf War and foretold the terrorist attacks of September 11, 2001."

Observe Orient Decide Act (OODA) Loop

- *"The OODA loop is the cycle observe–orient–decide–act, developed by military strategist and USAF Colonel John Boyd. Boyd applied the concept to the combat operations process, often at the operational level during military campaigns. It is now also often applied to understand commercial operations and learning processes. The approach explains how agility can overcome raw power in dealing with human opponents."* – Wikipedia
- All effective purposeful military activity can be conceived in terms of OODA loop feedback process, especially at tactical/operational levels.
- Aligning Ethical Control mission design with OODA loop ensures that unmanned systems understandably partner within human-run teams.

OODA Significance for Ethical Control

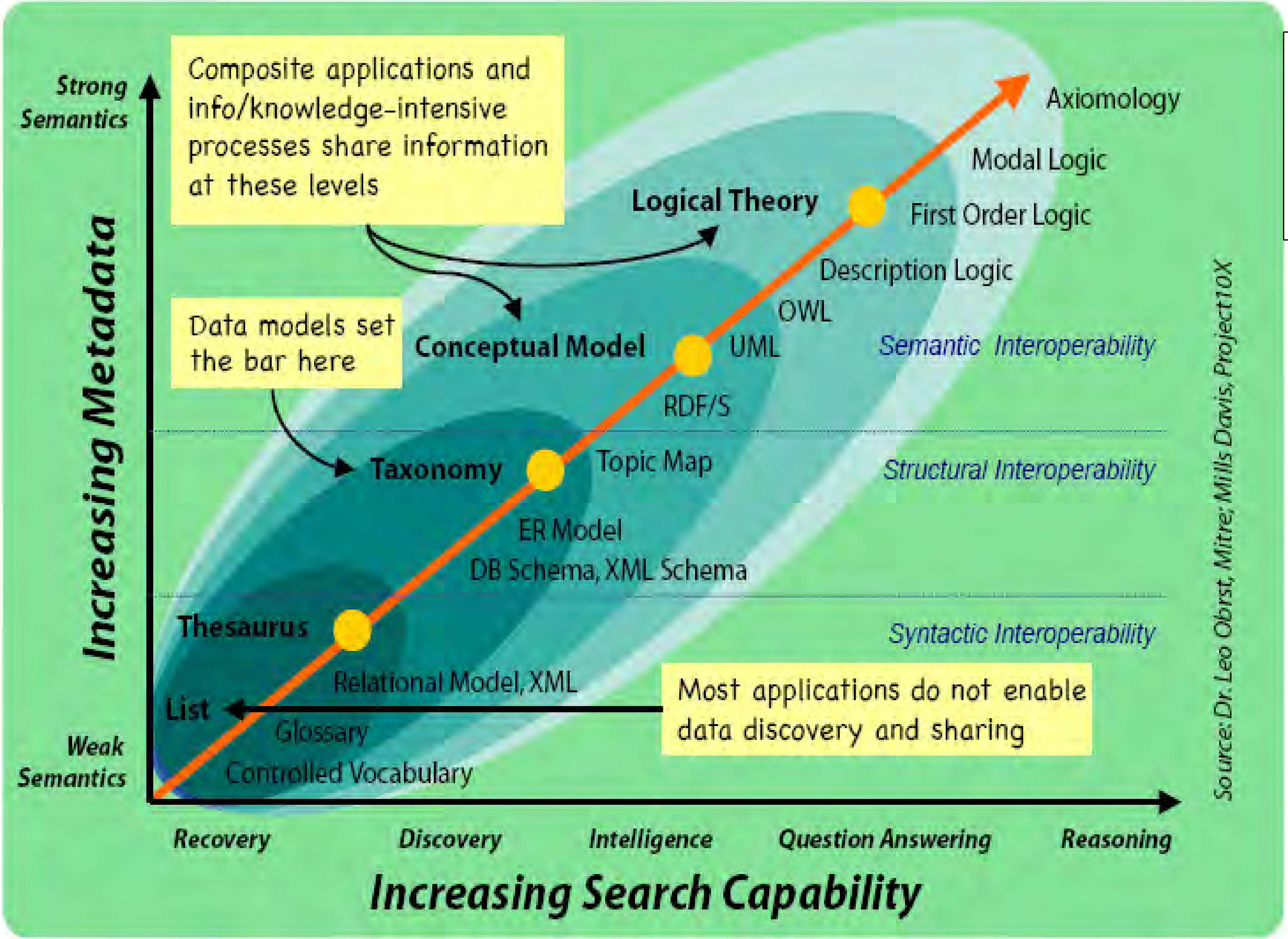
Classical robotic Sense-Decide-Act cycle for closed-loop control is insufficient for proper delegation of lethal (or lifesaving) force to unmanned systems.

Observe-Orient-Decide-Act (OODA) Loop is essential for coherent operations.

- *Observe* includes direct sensing and communication inputs.
- *Orientation* includes thorough Rules of Engagement (ROE) constraints and identification friend/foe/neutral/unknown (IFFNU) of all relevant contacts.
- *Decision* logic of unmanned system tactics, techniques, procedures (TTP) includes authorization and confirmation by human supervisors, either in real-time or in advance, for critical steps leading to lethal force.
- *Actions* in tandem with direct or intermittent human supervisory command enables effective Ethical Control of remote systems.

Feedback loops are essential, generally leading to… *more effective operations.*

Improving Semantic Representation



Semantic Web Stack

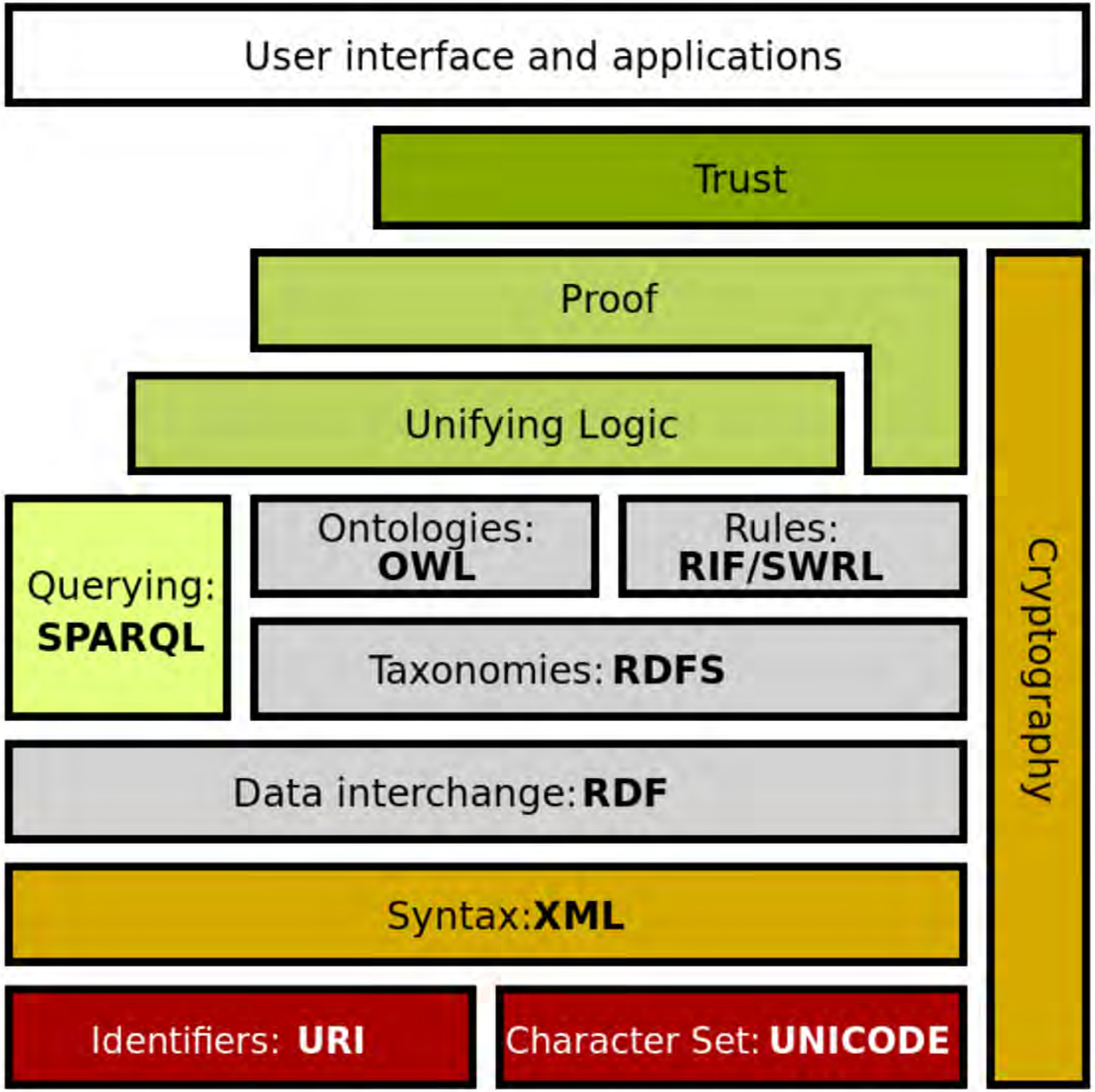
Extends larger Web architecture

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

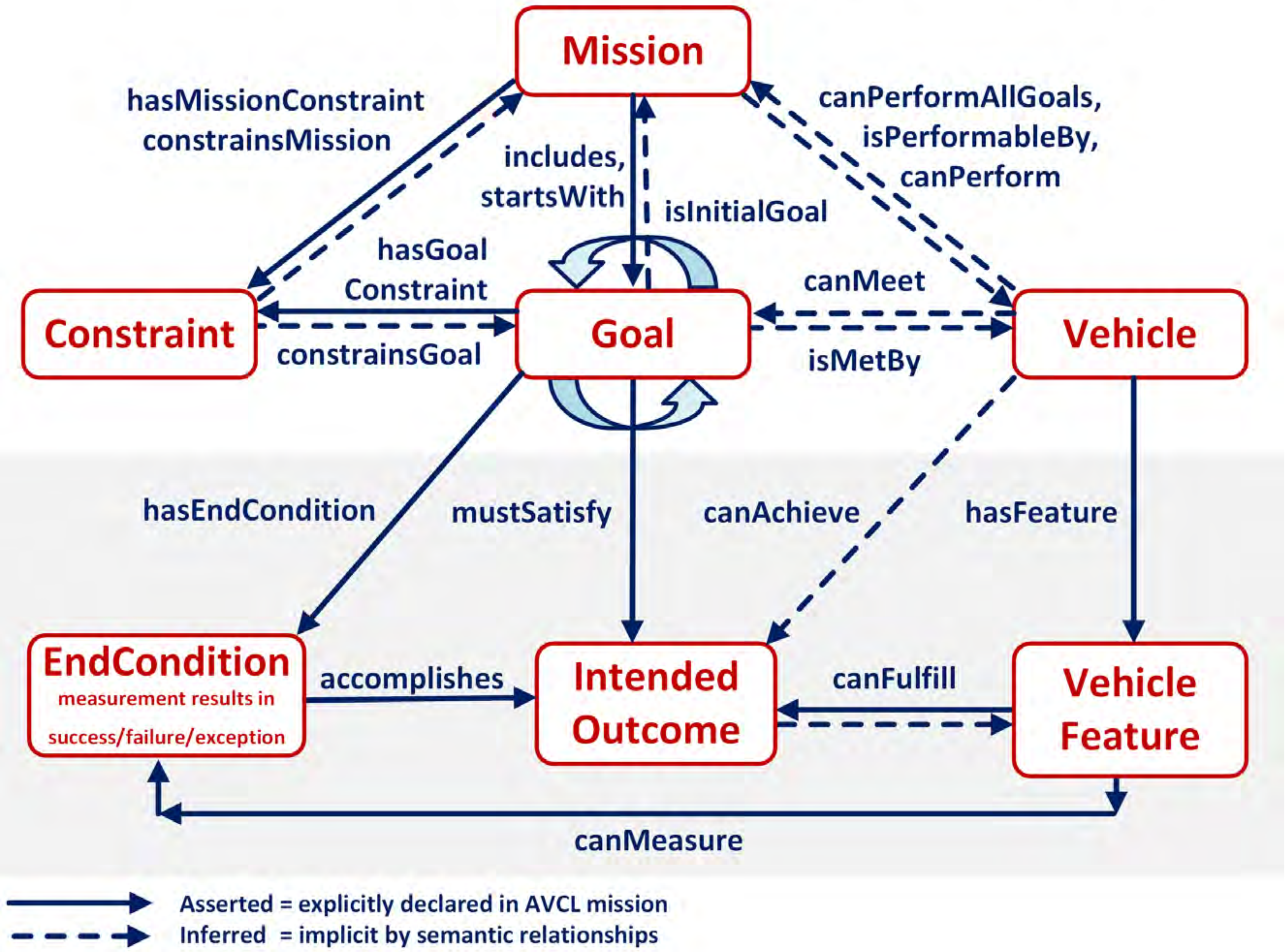
Trusting derived (composed) statements arises from

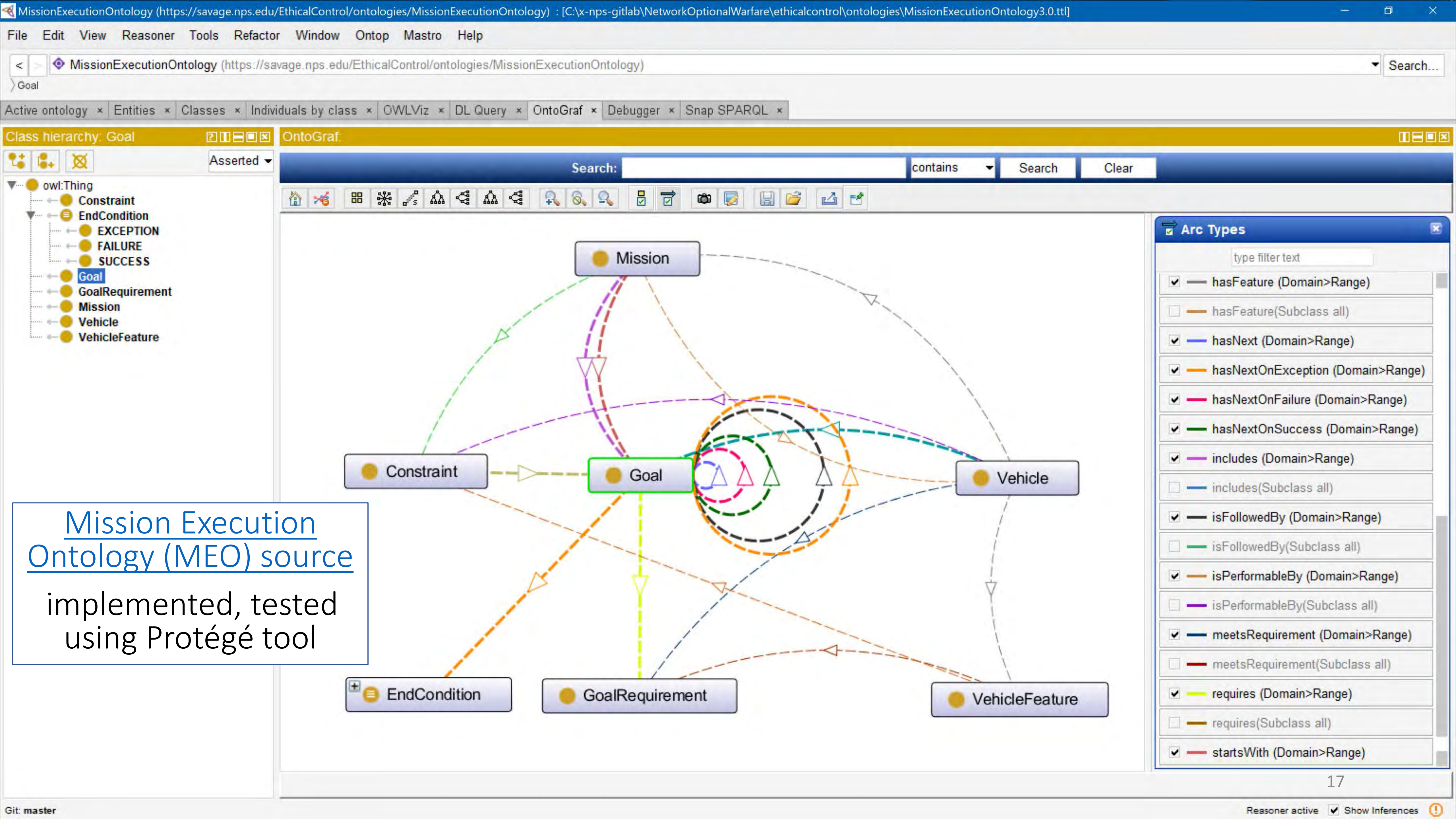
- Encryption + digital signature confirms trusted data sources
- Formal logic is basis for deriving new information
- Wikipedia: [Semantic Web Stack]

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



Mission Execution Ontology (MEO) 3.0





[Mission Execution Ontology (MEO) source]

implemented, tested using Protégé tool

```
@prefix :     <https://www.nps.edu/ontologies/MissionExecutionOntology> .
@prefix meo:  <https://www.nps.edu/ontologies/MissionExecutionOntology#> .
@prefix owl:  <http://www.w3.org/2002/07/owl#> .
@prefix rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xml:  <http://www.w3.org/XML/1998/namespace> .
@prefix xsd:  <http://www.w3.org/2001/XMLSchema#> .
@base <https://www.nps.edu/ontologies/MissionExecutionOntology> .

<https://www.nps.edu/ontologies/MissionExecutionOntology> rdf:type owl:Ontology .

#################################################################
#    Object Properties
#################################################################

###  https://www.nps.edu/ontologies/MissionExecutionOntology#appliesTo
meo:appliesTo rdf:type owl:ObjectProperty ;
  rdfs:domain meo:Constraint ;
  rdfs:range [ rdf:type owl:Class ;
               owl:unionOf ( meo:Goal meo:Mission ) ] ;
  owl:propertyChainAxiom ( meo:appliesTo meo:includes ) ;
  rdfs:comment "A Constraint applies to one or more Missions and/or one or more Goals." .
[ rdf:type owl:Axiom ;
  owl:annotatedSource meo:appliesTo ;
  owl:annotatedProperty rdfs:range ;
  owl:annotatedTarget [ rdf:type owl:Class ; owl:unionOf ( meo:Goal meo:Mission ) ] ;
  rdfs:comment "A Constraint can apply to a Mission or a Goal (and nothing else)." ;
  rdfs:label "C1" ] .
[ rdf:type owl:Axiom ;
  owl:annotatedSource meo:appliesTo ;
  owl:annotatedProperty owl:propertyChainAxiom ;
  owl:annotatedTarget ( meo:appliesTo meo:includes ) ;
  rdfs:comment "A Constraint that applies to a Mission must also apply to all of the Goals that Mission includes." ;
  rdfs:label "C3" ] .
```

[Mission Execution Ontology (MEO) source](https://www.nps.edu/ontologies/MissionExecutionOntology)

implemented, tested using Protégé tool

Turtle (.ttl) syntax

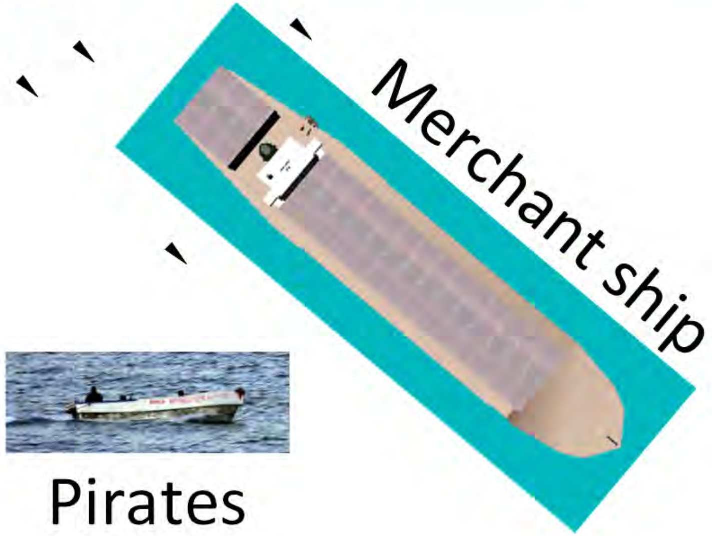
Life boat

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



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

Sailor Overboard Mission: Description

Purpose

- Life saving: single unmanned air/surface vehicle actions to complement human responses when performing *“SAILOR OVERBOARD”* operations.
- Carried out in direct concert with formal *shipboard emergency procedures*.
- Multiple UAVs/USVs might be employed in parallel with ships and aircraft, avoid mutual interference by each following deconflicted mission orders.

Phases

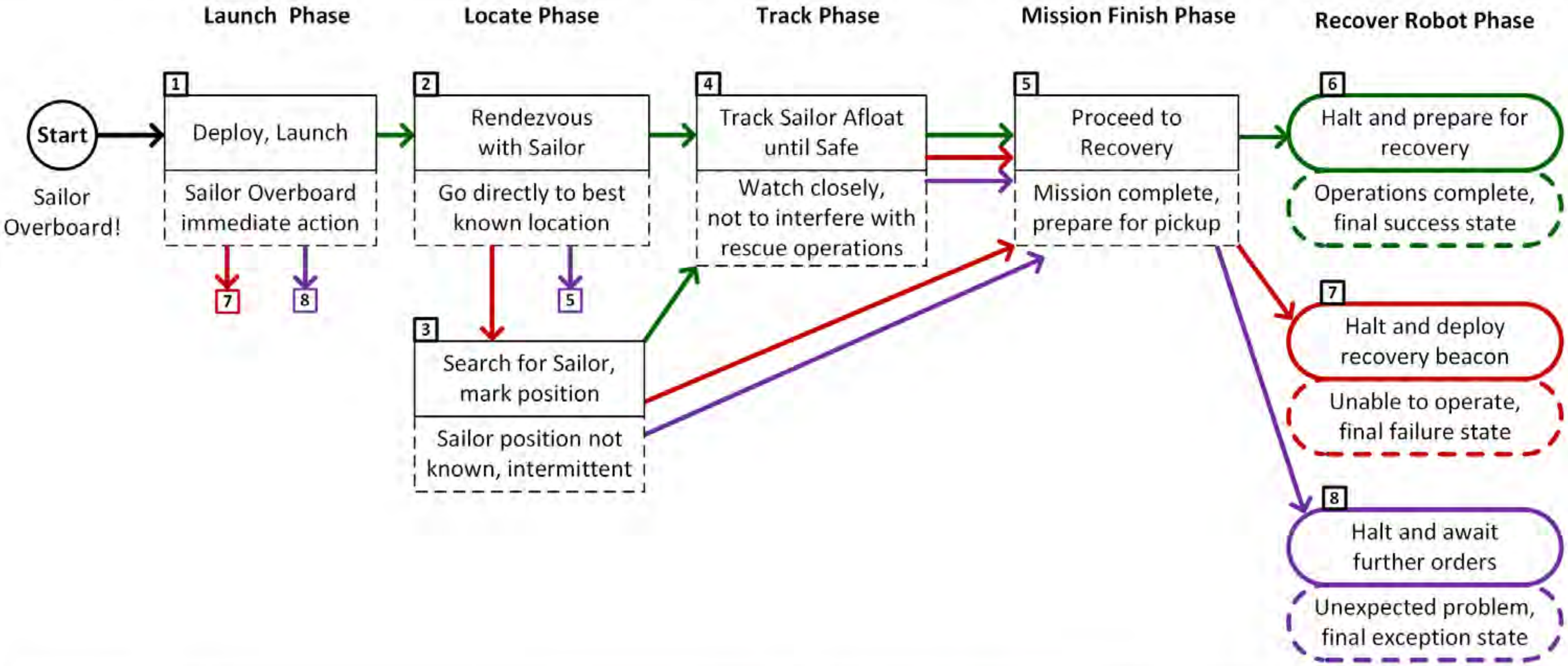
- Deploy/Launch, Rendezvous, Track Sailor until Safe, Return/Recovery.

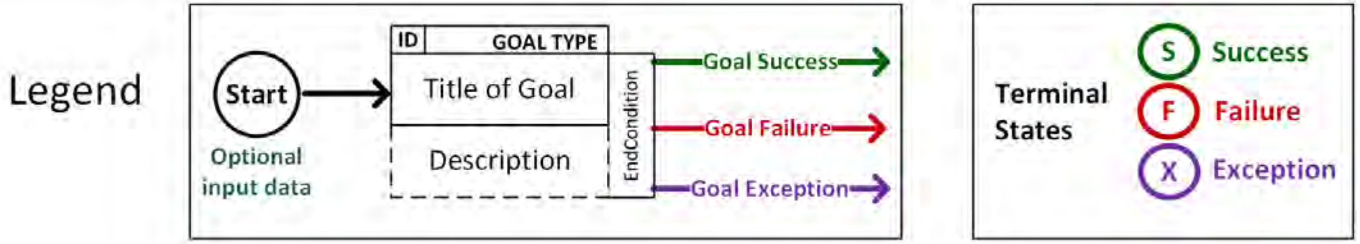
Human Supervisory Role

- Order standoff if interfering, manual control is possible due to proximity, can communicate to sailor via loudspeaker or beacon light.

Sailor Overboard, 8 Phases – Mission Execution Automaton (MEA)

Single unmanned air/surface vehicle actions to complement human response when performing "**SAILOR OVERBOARD**" operations, carried out in concert with **shipboard emergency procedures.** Multiple UAVs/USVs can be employed in parallel with ships/aircraft, each following mission orders.





Don Brutzman and Bob McGhee
Mission upgrade 19 NOV 2019

SailorOverboard.xml in gitlab.nps.edu version control

https://gitlab.nps.edu/Savage/EthicalControl/blob/master/missions/avcl/SailorOverboard.xml

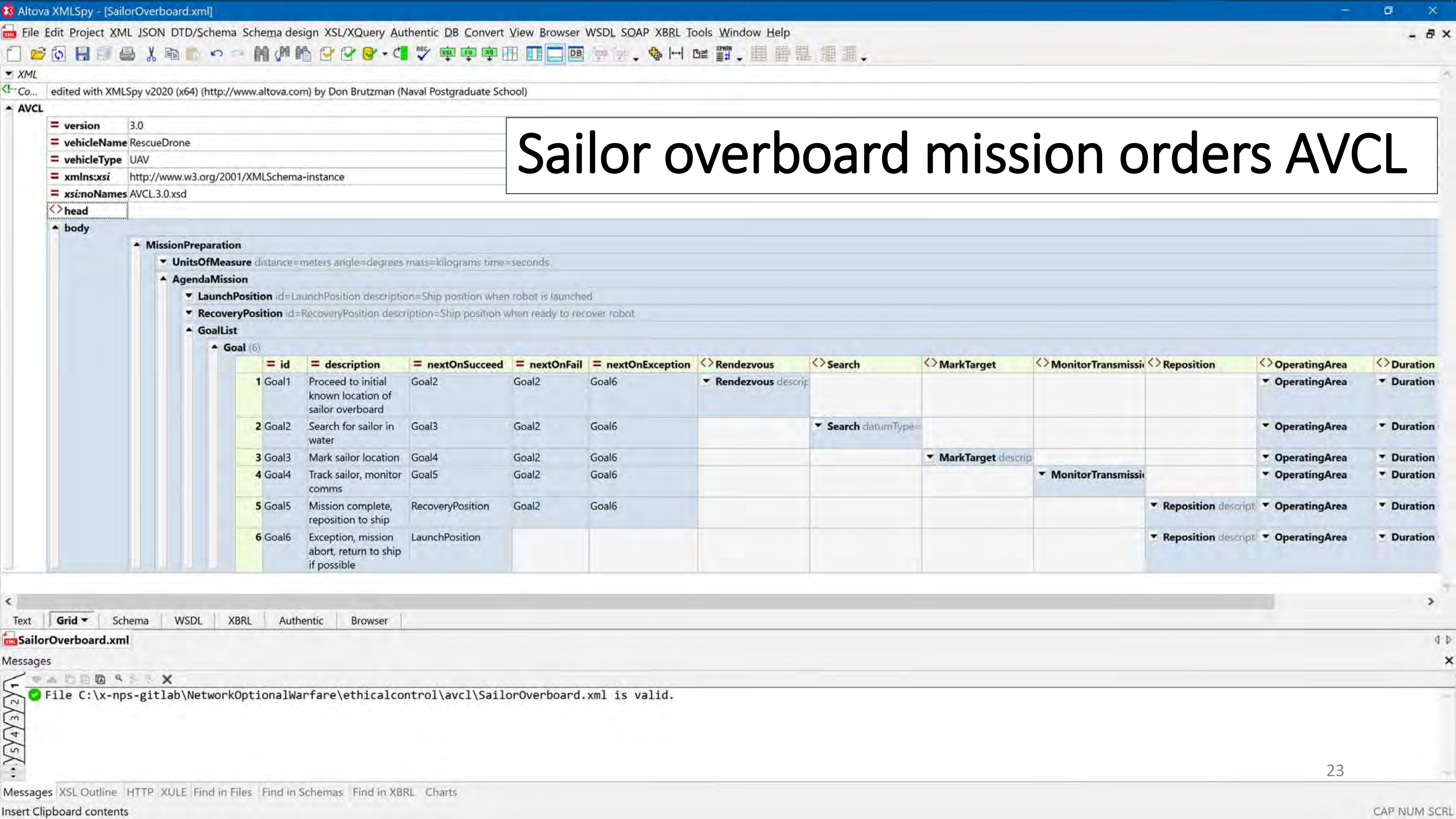
GitLab Projects Groups More

SailorOverboard.xml 5.63 KB

Edit Web IDE Replace Delete

```xml
<?xml version="1.0" encoding="UTF-8"?>
<AVCL version="3.0" vehicleName="RescueDrone" vehicleType="UAV" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="../../avcl/AVCL.3.0.xsd">
	<head>
		<meta name="title" content="SailorOverboard.xml"/>
		<meta name="created" content="19 November 2019"/>
		<meta name="modified" content="31 December 2019"/>
		<meta name="creator" content="Don Brutzman"/>
		<meta name="reference" content="https://wiki.nps.edu/display/NOW/Ethical+Control+of+Unmanned+Systems"/>
		<meta name="reference" content="https://gitlab.nps.edu/Savage/EthicalControl/raw/master/missions/SailorOverboard8PhaseMission.png"/>
		<meta name="identifier" content="https://gitlab.nps.edu/Savage/EthicalControl/tree/master/missions/avcl/SailorOverboard.xml"/>
		<meta name="generator" content="Altova XMLSpy, https://www.altova.com"/>
		<meta name="generator" content="Apache NetBeans, https://netbeans.apache.org"/>
		<meta name="generator" content="NPS Autonomous Unmanned Vehicle (AUV) Workbench, https://savage.nps.edu/AuvWorkbench"/>
		<meta name="license" content="../license.html"/>
	</head>
	<body>
		<MissionPreparation>
			<UnitsOfMeasure distance="meters" angle="degrees" mass="kilograms" time="seconds"/>
			<AgendaMission>
				<LaunchPosition id="LaunchPosition" description="Ship position when robot is launched">
					<LatitudeLongitude latitude="0" longitude="0"/>
				</LaunchPosition>
				<RecoveryPosition id="RecoveryPosition" description="Ship position when ready to recover robot">
					<LatitudeLongitude latitude="0" longitude="0"/>
				</RecoveryPosition>
				<GoalList>
					<Goal id="Goal1" title="Deploy, Launch" description="Sailor Overboard immediate Action" nextOnSucceed="Goal2" nextOnFail="Goal2" nextOnException="Goa
						<Rendezvous description="">
							<TargetVehicleID value="101" description="Sailor in water"/>
						</Rendezvous>
						<OperatingArea>
							<Point>
								<LatitudeLongitude latitude="36.62" longitude="121.506"/>
								<!--36°36'11"N 121°53'37"W-->
							</Point>
						</OperatingArea>
						<Duration value="300"/>
```

Sailor overboard mission orders AVCL



Ontology for Ethical Control of Unmanned Systems in a Surrogate Scenario: Example Relationship Definitions

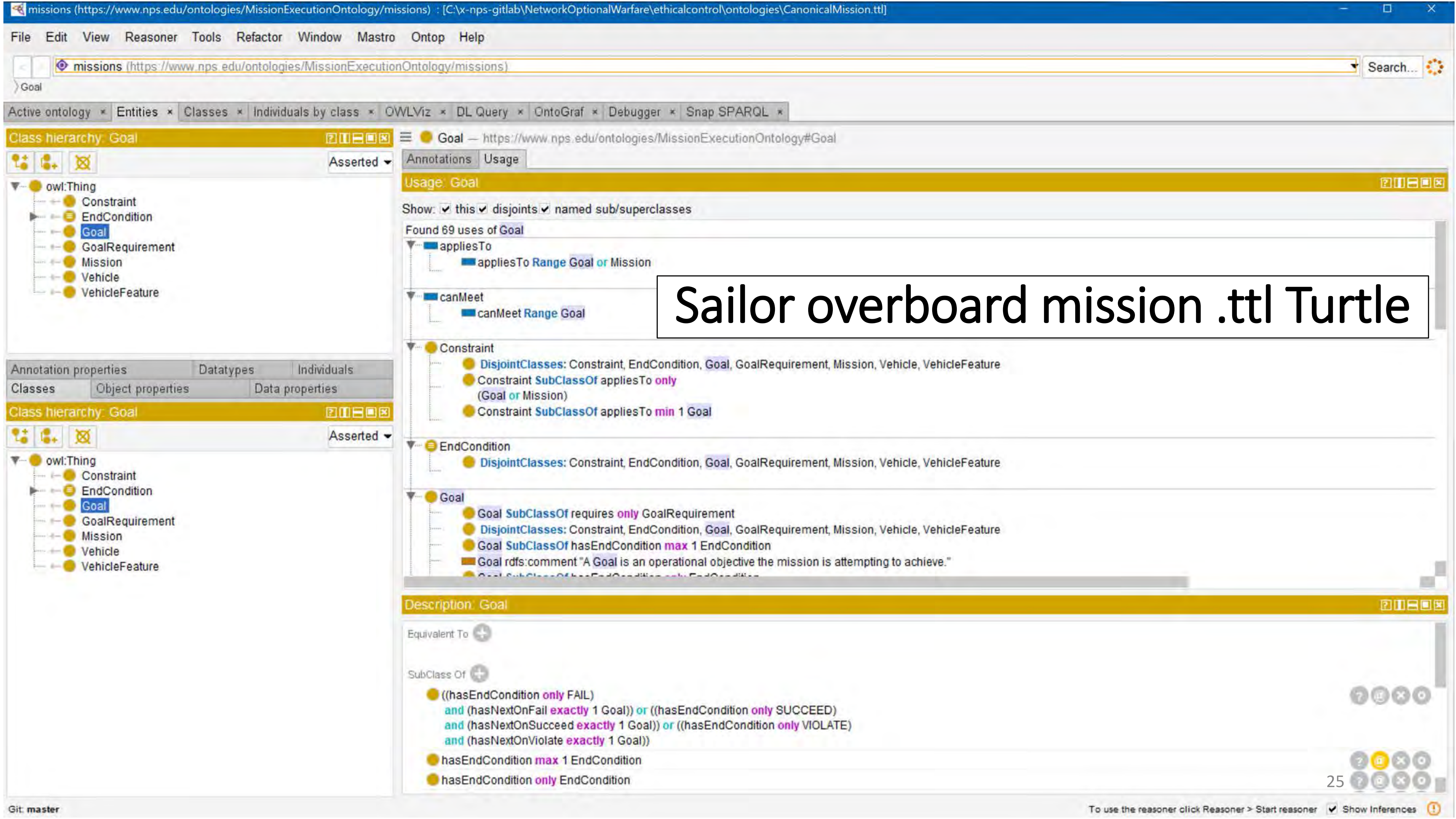
Mission Definition Expressed in Subject-Predicate-Object Triples Using Semantic Web Standards

Excerpt of Sailor Overboard Mission, expressed in Turtle Syntax:

```
###  https://www.nps.edu/ontologies/MissionExecutionOntology/missions#Goal_Launch
        :Goal_Launch rdf:type owl:NamedIndividual ;
                meo:hasNextOnFail :Goal_FailureDiagnosis ;
                meo:hasNextOnSucceed :Goal_TransitSearch ;
                meo:hasNextOnViolate :Goal_FailureDiagnosis .

###  https://www.nps.edu/ontologies/MissionExecutionOntology/missions#Goal_TransitSearch
        :Goal_TransitSearch rdf:type owl:NamedIndividual ;
               meo:hasNextOnFail :Goal_SearchForSailorAdrift ;
               meo:hasNextOnSucceed :Goal_TrackSailorAfloat ;
               meo:hasNextOnViolate :Goal_FailureDiagnosis .
```

Sailor overboard mission .ttl Turtle



first mission query MissionGoalsQuery_01 - hooray!

Brutzman, Don authored 1 hour ago

MissionGoalsQuery_01.rq 1.7 KB

```
PREFIX :     <https://www.nps.edu/ontologies/MissionExecutionOntology/missions#>
PREFIX meo:  <https://www.nps.edu/ontologies/MissionExecutionOntology#>
PREFIX owl:  <http://www.w3.org/2002/07/owl#>
PREFIX rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xml:  <http://www.w3.org/XML/1998/namespace>
PREFIX xsd:  <http://www.w3.org/2001/XMLSchema#>

# @base <https://www.nps.edu/ontologies/MissionExecutionOntology/missions>

# MissionGoalsQuery_01.rq    Query to list all goals with corresponding description information and sequencing logic.

################################################

SELECT ?goal ?nextOnSucceed ?nextOnFail ?nextOnViolate ?isPartOfPhase ?description # ?GoalFound ?phase

WHERE
{
    ?goal   a meo:Goal ; # Shorthand expression: a = rdf:type
            meo:isPartOfPhase       ?isPartOfPhase ; # TODO what about when no value is provided
            meo:hasNextOnSucceed    ?nextOnSucceed ;
            meo:hasNextOnFail       ?nextOnFail ;
            meo:hasNextOnViolate    ?nextOnViolate ;# TODO rename as Exception
            rdfs:comment            ?description .

    # https://stackoverflow.com/questions/11234371/sparql-query-results-without-namespace
    BIND (strafter(xsd:string(?goal),"#")              AS ?GoalFound)
    BIND (strafter(xsd:string(?nextOnSucceed),"#")     AS ?GoalNextOnSucceed)
    BIND (strafter(xsd:string(?nextOnFail),"#")        AS ?GoalNextOnFail)
    BIND (strafter(xsd:string(?nextOnViolate),"#")     AS ?GoalNextOnViolate)
#   BIND (coalesce(?isPartOfPhase,"")                  AS ?phase)
}
ORDER BY (?GoalFound) # alphanumeric order results in order given by each name
################################################
```

SPARQL mission query

[MissionQuery_01_GoalBranches.rq]

SPARQL query response

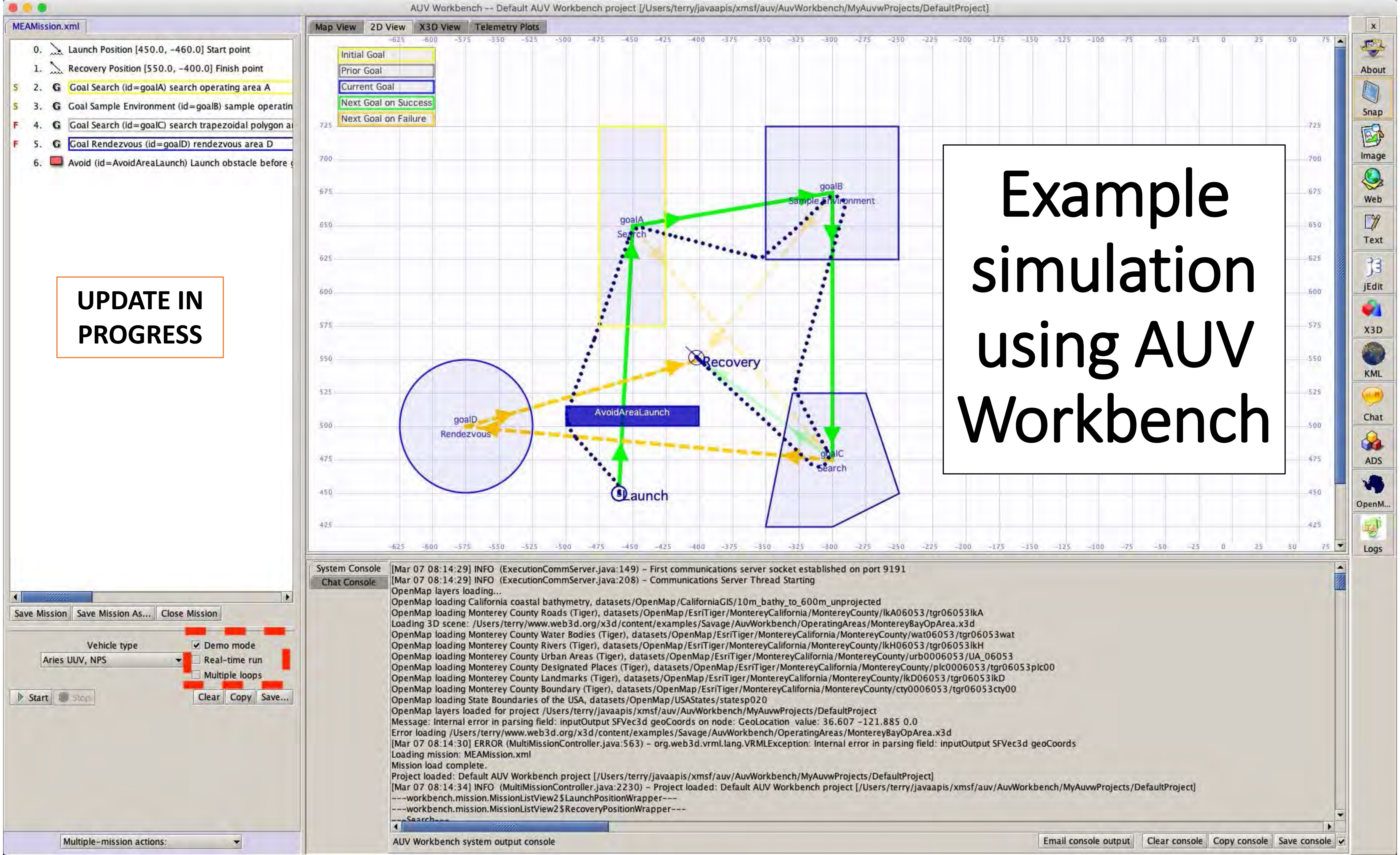
SailorOverboardConverted. MissionQuery_01_GoalBranches.rq.txt

goal	nextOnSuccess	nextOnFailure	nextOnException	isPartOfPhase	description
:Goal1	:Goal2	:Goal2	:Goal5	"Launch"	"Deploy, Launch: Sailor Overboard Immediate Action"
:Goal2	:Goal4	:Goal3	:Goal5	"Locate"	"Rendezvous with Sailor: Go directly to best known location"
:Goal3	:Goal4	:Goal5	:Goal5	"Locate"	"Search for Sailor: Sailor position not known, intermittent"
:Goal4	:Goal5	:Goal5	:Goal5	"Track"	"Track Sailor afloat until safe: Watch closely, not to interfere with rescue operations"
:Goal5	:Goal6	:Goal2	:Goal6	"Mission Finish"	"Proceed to Recovery: Mission complete, prepare for pickup"
:Goal6				"Recover Robot"	"Halt and prepare for recovery: Operations complete, final success state"
:Goal7				"Recover Robot"	"Halt and deploy recovery beacon: Unable to continue, final failure state"
:Goal8				"Recover Robot"	"Halt and await further orders: Unexpected problem, final exception state"

SPARQL query response

SailorOverboardConverted. MissionQuery_02_InitialGoal.rq.txt

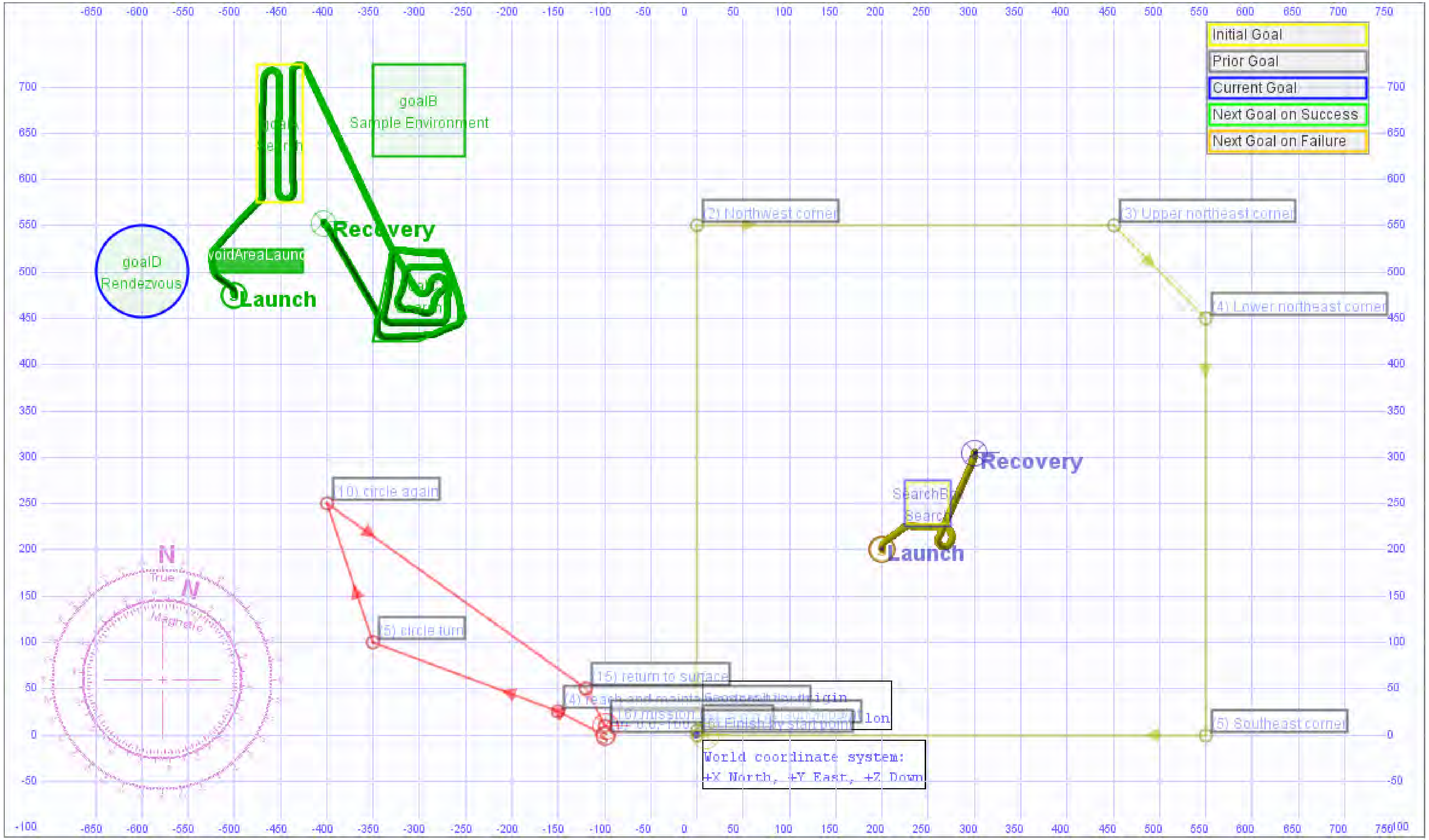
Mission	InitialGoal	isPartOfPhase	goalDescription
:SailorOverboard	:Goal1	"Launch"	"Deploy, Launch: Sailor Overboard Immediate Action"



UPDATE IN PROGRESS

Example simulation using AUV Workbench

4 earlier example missions, UUV and USV



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

Not suitable for brute-force numerical computation

- AI algorithms for Machine Learning (ML) and Data Mining are often based on statistically training against large datasets to find patterns for filters.
  - For example, **convolutional neural networks**, **genetic algorithms**, **reinforcement learning**, etc.
- Often requires identifying right/wrong matches within large search spaces.
- Such predictive analytics are useful for classification models using detailed and noisy sensor data. Given the central importance of IFFNU and some conditional communications to ethical control, ML filters can be helpful if carefully applied.
- Nevertheless such approaches are not appropriate for carefully following Rules of Engagement (ROEs), Laws of Armed Conflict (LOAC) or other ethical prerequisites, especially when human expertise and judgement is essential for robot teams.
- (Similarly, massive computation or Quantum Computing approaches might be useful in some problems, but are not of practical use for Ethical Control mission orders given by human commanders judiciously guiding remote mobile robots)

Naval history has long shown that sound human judgement is crucial for assessing best strategies and courses of action in ill-structured contexts. Semantic Web approaches are preferable and actionable for Ethical Control.

US Semantic Technologies Symposium (US2TS)
Session: [Hybrid AI for Context Understanding]

US2TS
10-12 MAR 2020

What is nature of context that described hybrid AI system is trying to understand?

- Validate human orders to remote unmanned systems with capacity for lethal/lifesaving force.
- Ethics for Rules of Engagement (ROE), Laws of Armed Conflict (LOAC), operational constraints.

What specific methods and technologies does this hybrid AI system use, and how?

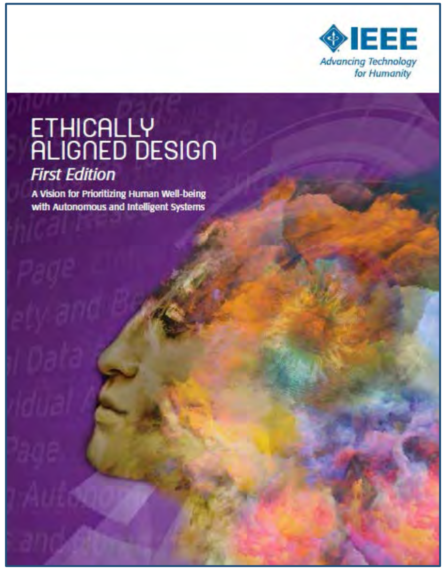
- Validatable XML mission syntax using controlled vocabularies with corresponding ontology.
- Perform SPARQL queries of RDF/OWL to check complex relationships, requirements, violations.
- Conversions for declarative orders, language implementations, Turtle triples with Protégé, ARQ.

What are current limitations in presented solution, what is plan for future work?

- Representative test cases are being tested in simulation to build out verification framework.
- Supports maritime operations with tractable Identification Friend Foe Neutral Unknown (IFFNU).
- Expand scope with larger mission sets for diverse operations by unmanned systems in real world.
- Bridge command and control (C2) with modeling and simulation (M&S) virtual environments, for
- Domain-expert qualification of hardware/software systems using tactical scenarios as unit tests.

IEEE P7000-series Standards Projects

https://ethicsinaction.ieee.org



- P7000 Model Process for Addressing Ethical Concerns during System Design
- P7001 Transparency of Autonomous Systems
- P7002 Data Privacy Process
- P7003 Algorithmic Bias Considerations
- P7004 Standard on Child and Student Data Governance
- P7005 Standard on Employee Data Governance
- P7006 Standard on Personal Data AI Agent Working Group
- P7007 Ontological Standard for Ethically driven Robotics and Automation Systems
- P7008 Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems
- P7009 Standard for Fail-Safe Design of Autonomous, Semi-Autonomous Systems
- P7010 Well-being metrics Standard for Ethical Artificial Intelligence and Autonomous Systems
- P7011 Stadard for the Porcess of Identifying and Rating the Trustworthiness of News Sources
- P7012 Standard for Machine Readable Personal Privacy Terms
- P7014 Standard for Ethical Considerations in Emulated Empathy in Autonomous and Intelligent Systems

IEEE Standards Project P7007 for Ontological Standard for Ethically driven Robotics and Automation Systems

- IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems.
  - https://ethicsinaction.ieee.org includes large document providing broad rationale.
  - Includes 15 separate working groups in IEEE Standards Association (IEEE-SA).
- Relevant group: P7007, Ethically driven Robotics and Automation Systems.
  - "IEEE P7007 Standards Project for Ontological Standard for Ethically driven Robotics and Automation Systems establishes a set of ontologies with different abstraction levels that contain concepts, definitions and axioms that are necessary to establish ethically driven methodologies for the design of Robots and Automation Systems."
  - http://standards.ieee.org/develop/project/7007.html
  - Must be IEEE member, observe patent-policy requirements to participate in working group.
  - "Not the intent to specify required ethical behaviors, but rather to formalize a vocabulary of terms, concepts, and relationships that can be used to enable unambiguous discussion among […] communities regarding what it means for autonomous systems to exhibit ethical behaviors."
  - Excellent forum with rich references, worth observation and participation.
- ***Active work:*** align several Ethical Control terms, concepts, use cases with P7007.

Unmanned Maritime Autonomy Architecture (UMAA)

Richard R. Burgess, "Navy Requests Information for Unmanned Maritime Autonomy Architecture," SEA POWER, 20 FEB 19

- "The intent of UMAA is to provide overarching standards that various UUVs and USVs can be built to in order to avoid creating multiple conflicting systems in the future"
- "The UMAA is being established to enable autonomy commonality and reduce acquisition costs across both surface and undersea unmanned vehicles."
- Topics of interest include Situational Awareness, Sensor and Effector Management, Processing Management, Communications Management, Vehicle Maneuver Management, Vehicle Engineering Management, Vehicle Computing Management, Support Operations

Multiple public NAVSEA documents refer to autonomy efforts and UMAA.

- NAVSEA PMS 406 is Program Office for Unmanned Maritime Systems
- NAVSEA Fact Sheet: Unmanned Maritime Systems Program Office (PMS 406) https://www.navsea.navy.mil/Portals/103/Documents/Exhibits/SNA2019/UnmannedMaritimeSys-Small.pdf
- Automated Management of Maritime Navigation Safety
  Navy SBIR 2020.1 - Topic N201-059, https://www.navysbir.com/n20_1/N201-059.htm

Proposed Critical Path Forward

Acting SECNAV Modly: unmanned, data, wargame, iterate

- [SECNAV Modly: Path to 355 Ships Will Rely on New Classes of Warships]
- *USNI News*, Megan Eckstein, 3 February 2020

USNI News

"You look at the frigate program: we think, because of the way we've approached that program, we've probably taken three years off the product development lifecycle for that. So we have to start doing the same type of thing: looking at proven hulls, things that can be adaptable for different areas. I understand the Hill's concerns about ***unmanned***, and we get that. … We have to convince them with ***data***: we have to ***wargame*** this, we have to ***iterate*** it over and over again."

Honorable Thomas Modly, Acting SECNAV

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

Assessment of Current Thinking

- 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.

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

Contact

Curt Blais

clblais@nps.edu

home page

Code MV/Bl, Naval Postgraduate School

Monterey California 93943-5000 USA

1.831.656.3215 work

CRUSER · NEWS

Consortium for Robotics and Unmanned Systems Education and Research

NPS

PRAESTANTIA PER SCIENTIAM

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