Doctrine and Ethics-Compliant Autonomy using an Ontological Framework

for Ethical Control of Unmanned Systems

Workshop on Ethical Computing, AAAI Spring Symposium 2022

Metrics for Measuring AI's Proficiency and Competency for Ethical Reasoning

Don Brutzman, Curtis Blais, Sinker Wu, Rich Markeloff and Carl Anderson

Naval Postgraduate School (NPS), Raytheon Corporation

22 March 2022

Acknowledgements

This paper was ably distilled by Carl Anderson based on a project report by Don Brutzman, Curt Blais and Sinker Wu. Collaborative work is performed under a cooperative Research and Development Agreement (CRADA) between NPS and Raytheon. No dissemination restrictions exist for any work on Ethical Control of Unmanned Systems.

We thank Robert B. McGhee, Duane T. Davis, colleagues and numerous students for fundamental contributions over the past 30 years.

We dedicate this paper to the memory of Dr. Richard Markeloff with thanks for his many guiding insights.

Abstract

Ensuring ethical robot behavior requires complex representations and methodologies designed to guarantee it. Our approach extends frameworks already used by the U.S. military to ensure human ethical and doctrinal behavior by human beings. These have built in advantages of being able to express complex plans and constraints, yet remaining intelligible to humans, a requirement for ethical responsibility and liability.

To extend the framework to machines, mission constructs are expressed using an Autonomous Vehicle Command Language (AVCL) expressing mission actions and outcomes that can readily be translated to runnable source code in several programming languages. Missions written in AVCL can be validated via translation to an RDF/OWL Mission Execution Ontology (MEO) supporting queried proofs of ethical correctness. MEO ensures that missions are both semantically valid and compliant with ethical constraints. These technologies implement a simulation, testing, and certification regime that can serve as a foundation for human authority over and trust in robots capable of lethal force.

Motivation and concepts

What are we talking about?

Military Challenges for Ethically Compliant Systems

We are working within a military context, but applicability seems broad.

- Humans are qualified with legal, moral and command authority to carry out missions with potentially lethal (or indeed life-saving) force.
- Human teams have well-defined community ethos and doctrine, with tactics, techniques and procedures (TTPs) for precise communication and distributed execution on a global scale.
- Principles of International Humanitarian Law (IHL) and Laws of Armed Conflict (LOAC) are well understood by practitioners.
  - Perhaps not in legalistic terms, but absolutely understood in moral terms regarding whether mission directions are based on lawful orders.

People first – machines under human control

Only humans have authority, responsibility and accountability for actions with potential for lethal outcomes.

- Includes responsibility for unmanned systems under their command

Guiding heuristics for designing new AI-enabled mission capabilities:

a. How do humans accomplish such goal tasks today?
b. How might unmanned systems accomplish similar tasks in future?
c. How can human commanders safely direct and supervise such systems, retaining moral and legal authority over operations?

Criterion: Predictability

Robot control and associated development practices must be sufficiently reliable to predict robot behavior in any situation.

Anything less is relinquishing authority and responsibility.

- How do we measure that?
- How do we repeatably measure that?
- How to certify robots as competent to do what they are told, and not do what they are forbidden to do?

Criteria: Authority and Responsibility

Robot control must support ultimate (direct or indirect) control by qualified, well-informed humans over robot activities and outcomes.

- Dull dirty dangerous: actions may occur in hazardous environments, with intermittent communications, at lengthy *distances* and *durations.*

Because only human beings are accountable for moral responsibility, any robotic failures must be traceable back to a specific human entity (programmers, manufacturers, policy, operators, leadership, others).

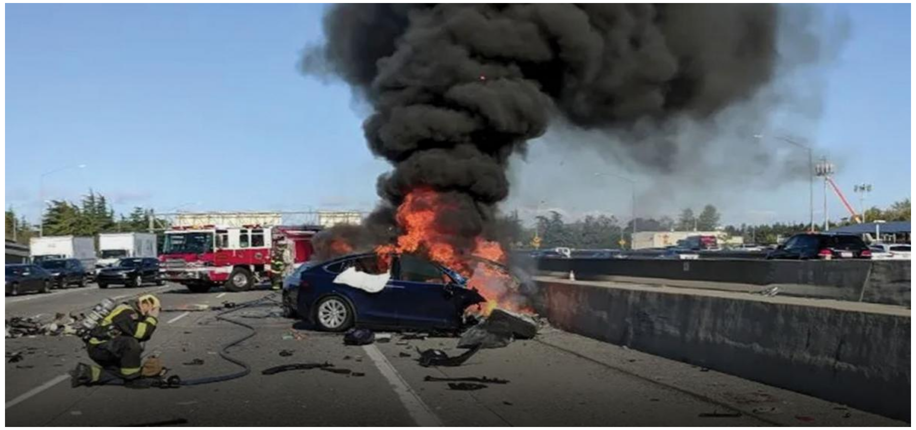
- Like human failures, everything subject to strict oversight and review

Artificial Intelligence: Too Fragile to Fight?

- **U.S. Naval Institute (USNI) PROCEEDINGS**
- *Automation—including AI—has persistent, critical vulnerabilities that must be thoroughly understood and adequately addressed if defense applications are to remain resilient and effective.*
- **Commander Edgar Jatho USN and Joshua A. Kroll, NPS**
- **February 2022, Vol. 148/2/1,428**

Machine learning (ML) and especially modern "deep learning" methods—the very methods driving the advances that make AI an important focus area today—are distinctly vulnerable to deception and perturbation.[3] Often, human-machine teaming is thought to be the solution to these issues, but such teaming itself is fraught and unexpectedly fragile in persistently problematic and counterintuitive ways

https://www.usni.org/magazines/proceedings/2022/february/artificial-intelligence-too-fragile-fight



Fatal crash of self-driving car, California 2018.

Criteria: Liability, Accountability, Culpability…

Liability assignment (whether legal or moral) requires that parties involved in robotic development and employment can reasonably foresee outcomes for which they are responsible.

- No really. This is due diligence. If in doubt, consider alternative…
- Currently hard to establish thresholds for acceptance, metrics needed.
- *Riddle:* Shiny new robot arrives, very expensive and powerful, guidance says “push big red AI button to operate.” *What to call that?*
- *Riddle:* Autonomous passenger vehicle confronted by trolley problem, must turn right/left or slam ahead or slam behind, fatalities likely.

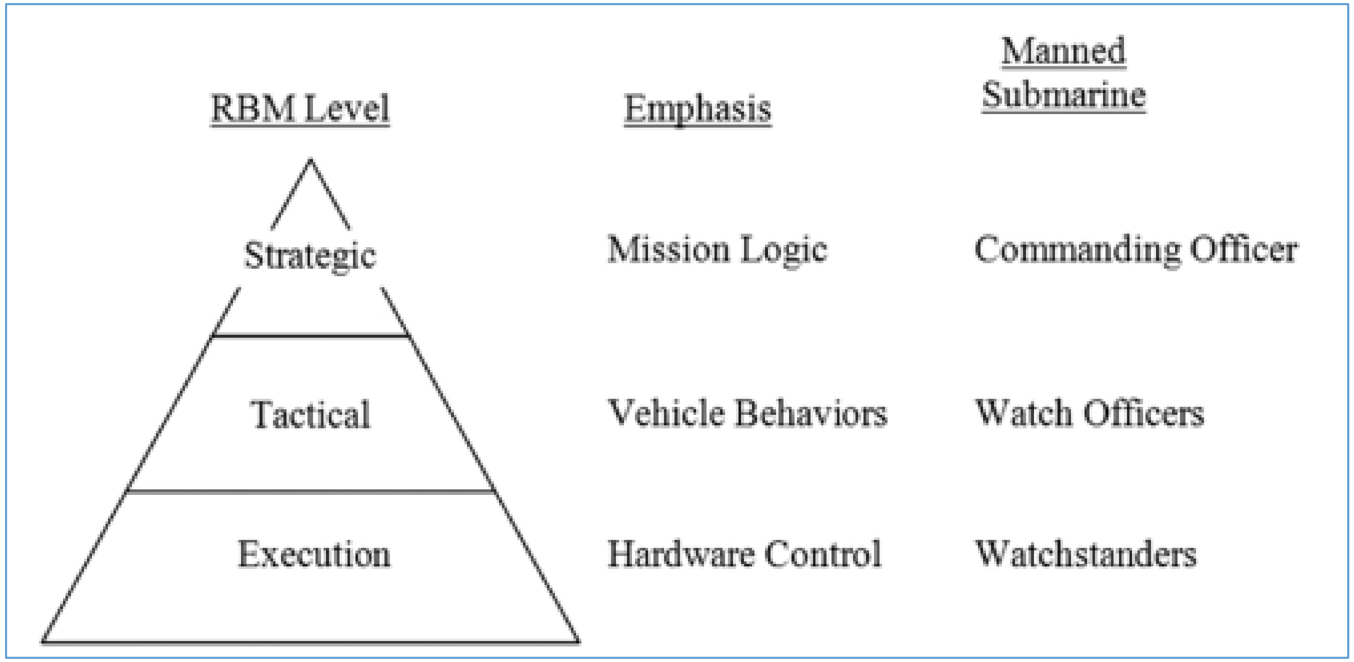
  What do we call that decision point in the software?

Mission-oriented approach, common across all platforms

Ethical operations require trust at all levels, up down left right

Foundational concept: Rational Behavior Model

- Three-level architectures provide good division of abstractions that often maps well to division of responsibility for human activities.
- We often see three-level architectures in all manner of robot designs



How to exploit similarities?

Treat robots as subordinates within the same framework. If not competent and trusted, then human orders are out of scope and unsound.

Foundational concept: communications are key

Human military actions are governed by formal **mission orders**

- Well-defined lexicons with varying terms, consistent semantics
- Structured orders follow guidance, who what when where how why
- Effective across diverse communities, countries (e.g. NATO procedures)
- Both operators and systems require coherent command language, otherwise orders are syntactically invalid or semantically confounded.

Syntactic validation with schema: is mission tasking precisely defined?

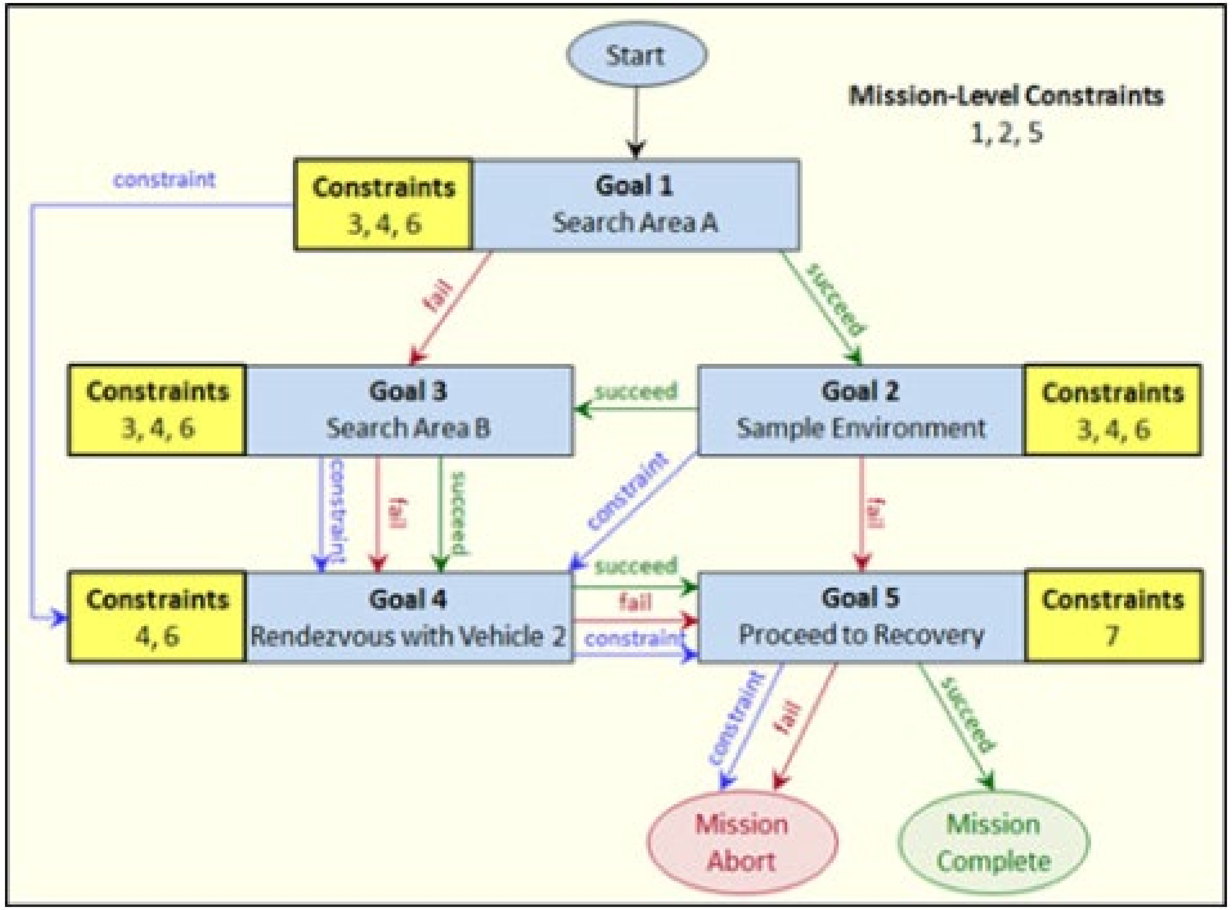
Semantic validation with ontology: is mission coherent, consistent, complete? Does it meet guidance? Does it avoid contradictions/omissions?

In combination: better AI support for human commanders, also testable.

Ternary goal logic: success, failure, exception

Conduct of mission orders can be expressed as series of goals connected as tree describing how to follow completion of each goal.

Original basis was binary decision tree. Ternary approach is more concise.



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

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

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.

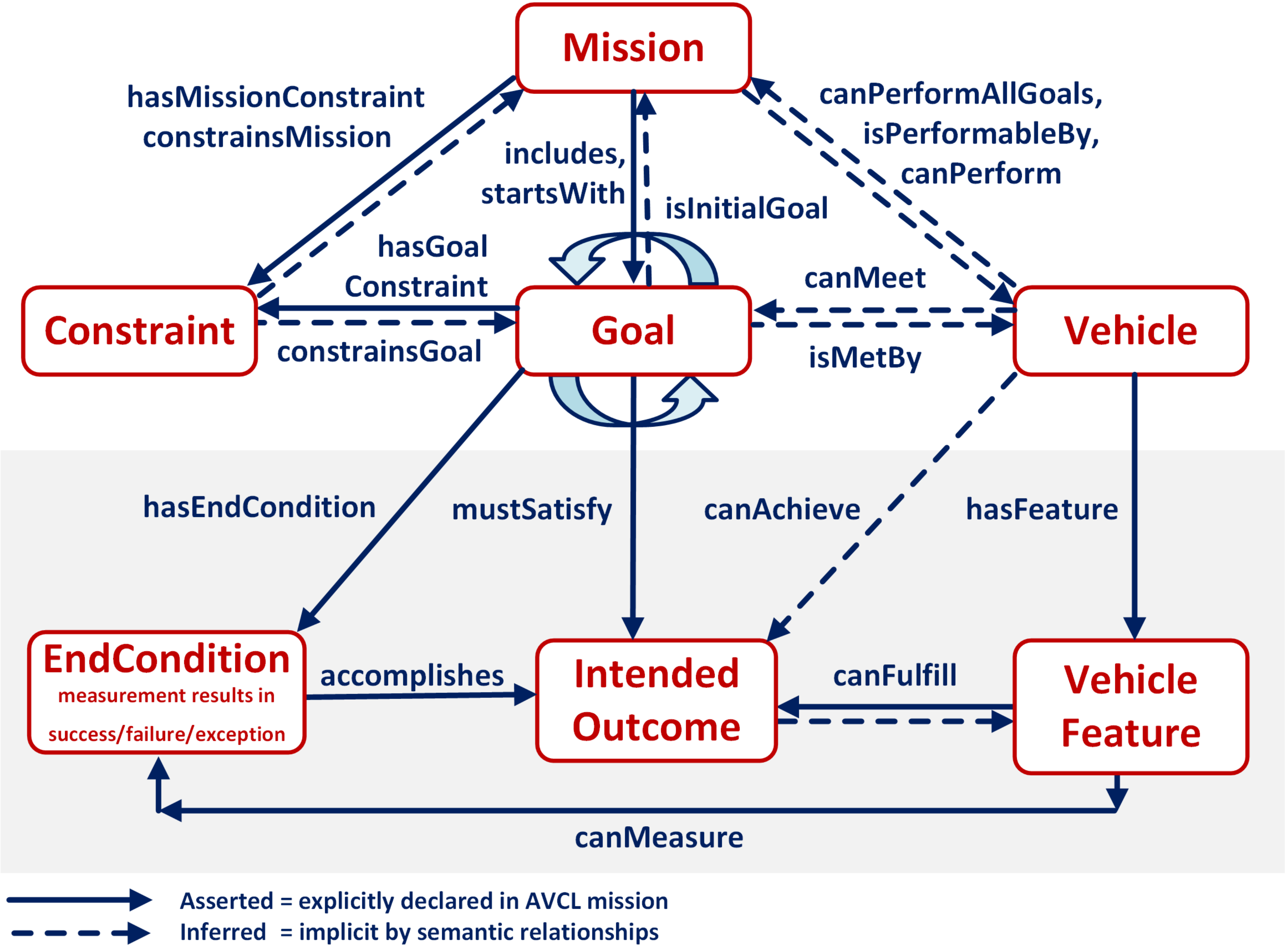
Description Logic (DL) Rules provide basis for Mission Execution Ontology (MEO)

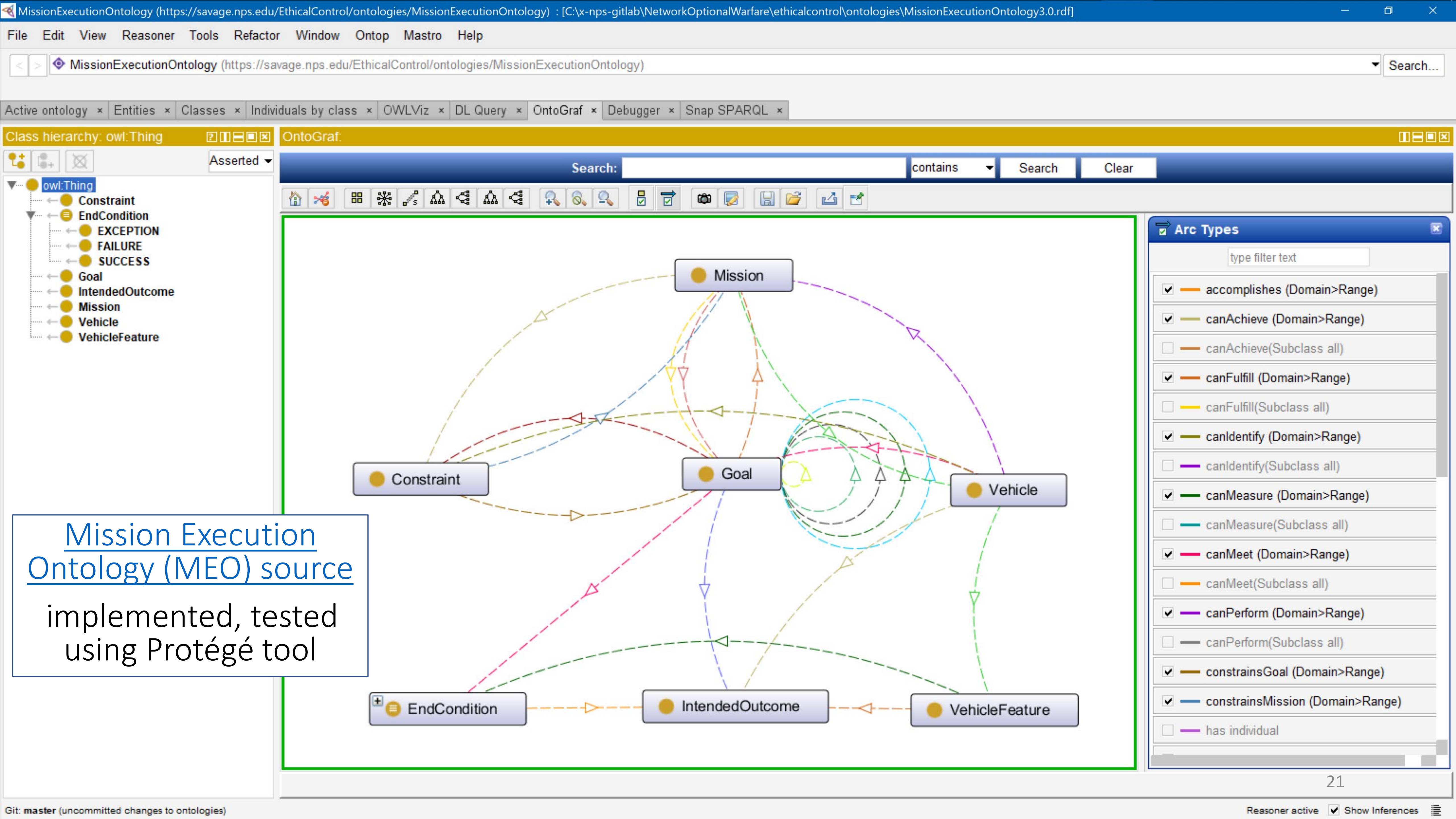
Rules	Description Logic Equations	Plain-language description
M = Mission Rules		
M1	Mission $\sqsubseteq$ $\forall$startsWith.Goal $\sqcap$ =1startsWith.Goal	A Mission can only start with a Goal and must start with exactly one Goal
M2	Mission $\sqsubseteq$ $\forall$includes.Goal $\sqcap$ $\geq$1includes.Goal	A Mission can only include Goals and must include one of more Goals
M3	Mission $\sqsubseteq$ $\forall$hasConstraint.Constraint	A Mission can only be constrained by Constraints
M4	startsWith $\sqsubseteq$ includes	A Mission must include the Goal that it starts with
M5	Mission $\sqsubseteq$ $\forall$performableBy.Vehicle	A Mission can only be performed by a Vehicle
M6	Cannot be expressed in DL	A Mission cannot be performable by a Vehicle unless that Vehicle has the ability to identify all Constraints associated with that mission
M7	Cannot be expressed in DL	A Mission cannot be performable by a Vehicle unless that Vehicle has the capability to accomplish all Goals included in that Mission

Excerpted from full
[Mission Execution Ontology Decision Logic Tables]

Original author: Duane Davis
Model updated by Curtis Blais in collaboration with Raytheon engineers through NPS-Raytheon CRADA

Mission Execution Ontology (MEO) 3.0





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

implemented, tested using Protégé tool

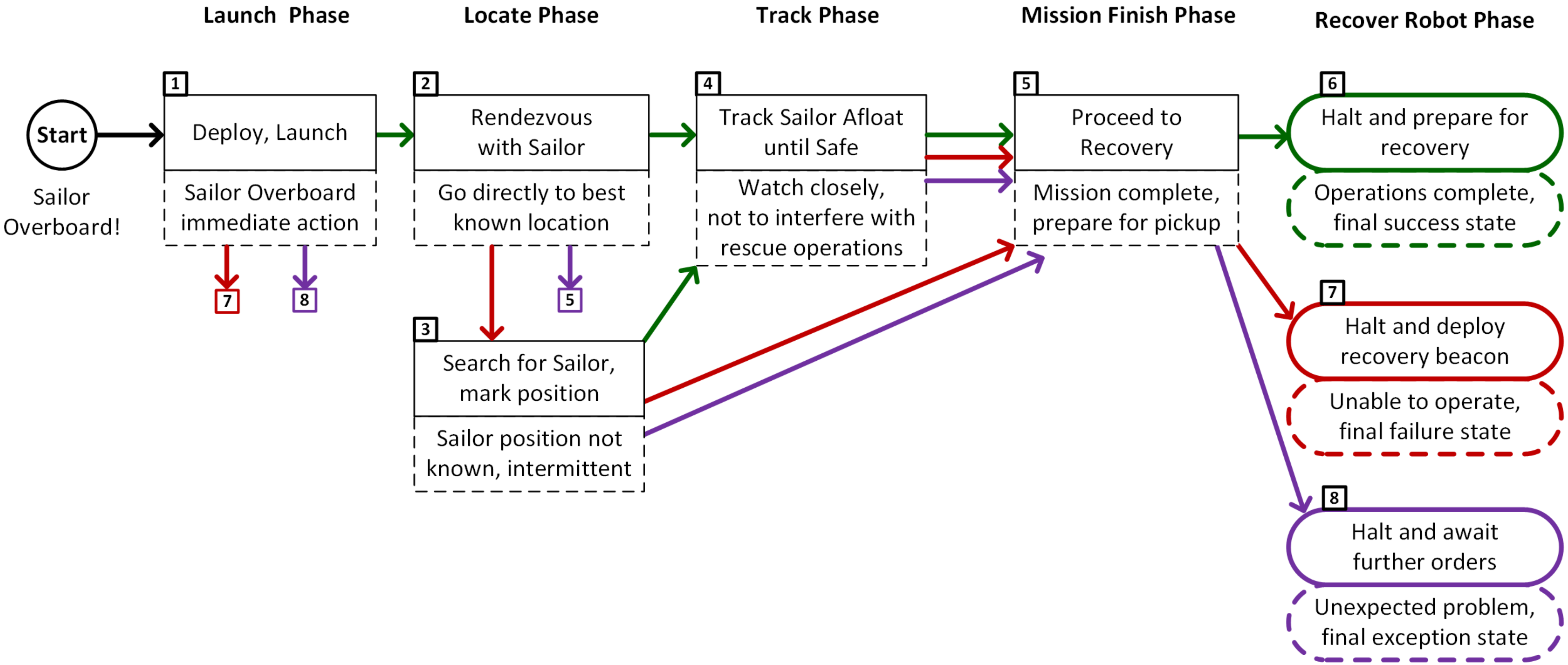
Variations on a theme: Sailor Overboard

Every sailor on board knows what to do. Will our robots perform well with others?

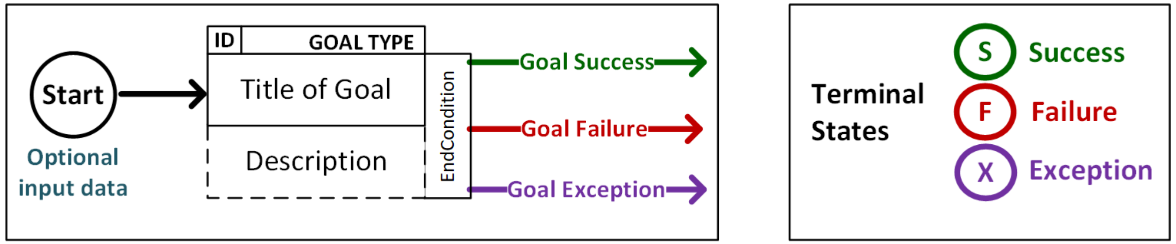
Buckle your safety belts, we will move fast!

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.

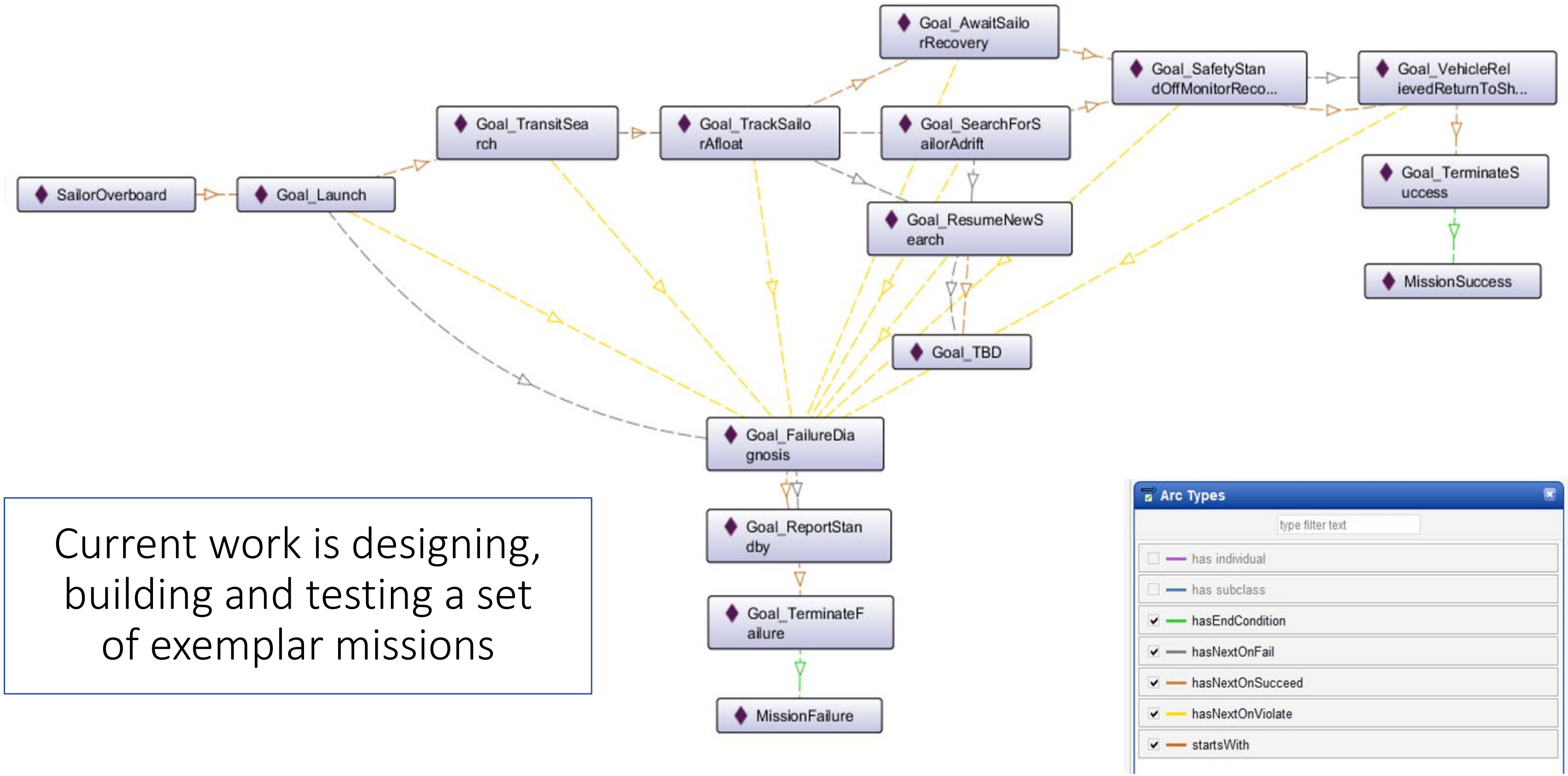


Legend



Don Brutzman and Bob McGhee
Mission upgrade 19 NOV 2019

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

SailorOverboard.xml in gitlab.nps.edu version control

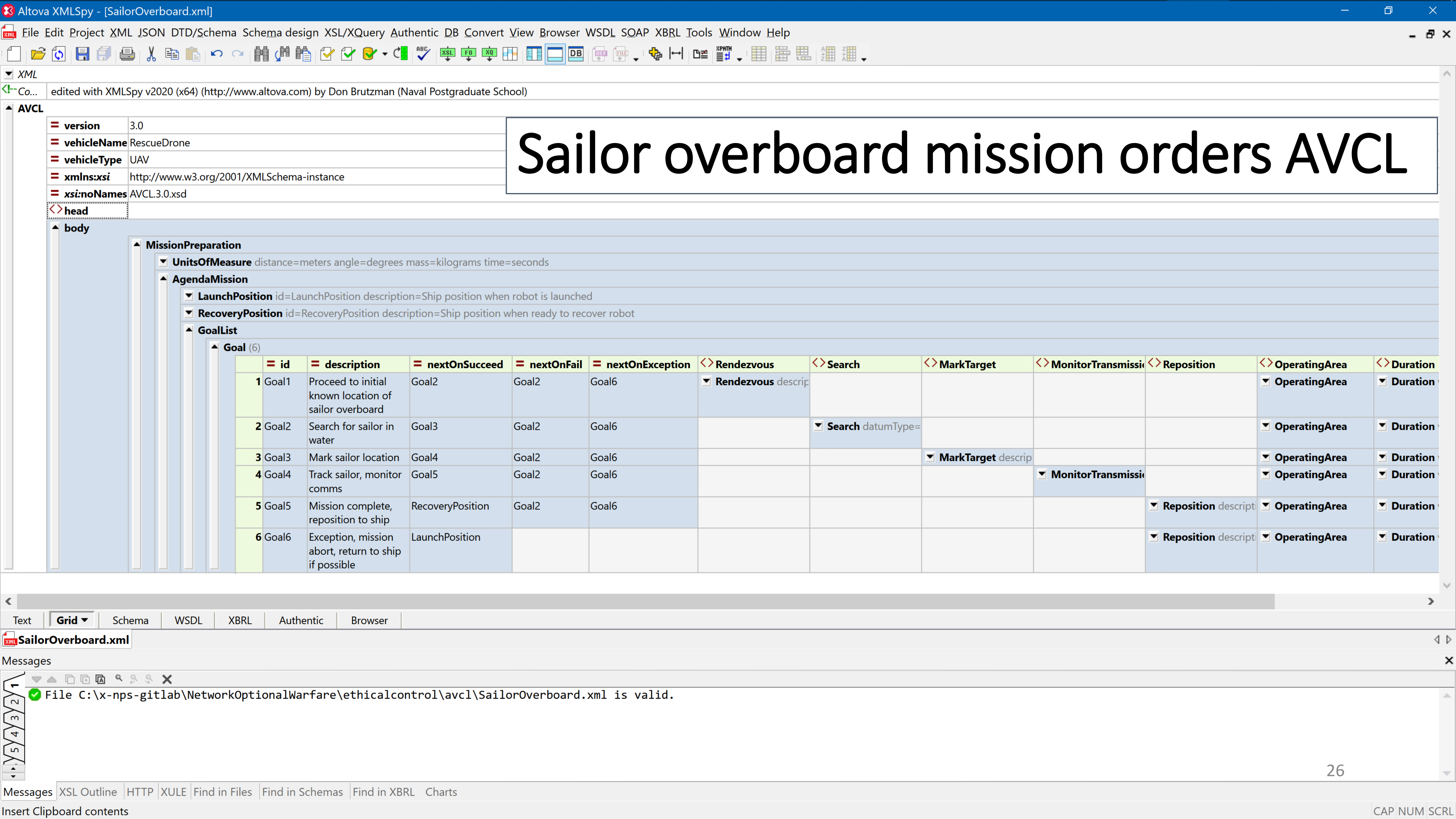
https://gitlab.nps.edu/Savage/EthicalControl/blob/master/missions/avcl/SailorOverboard.xml

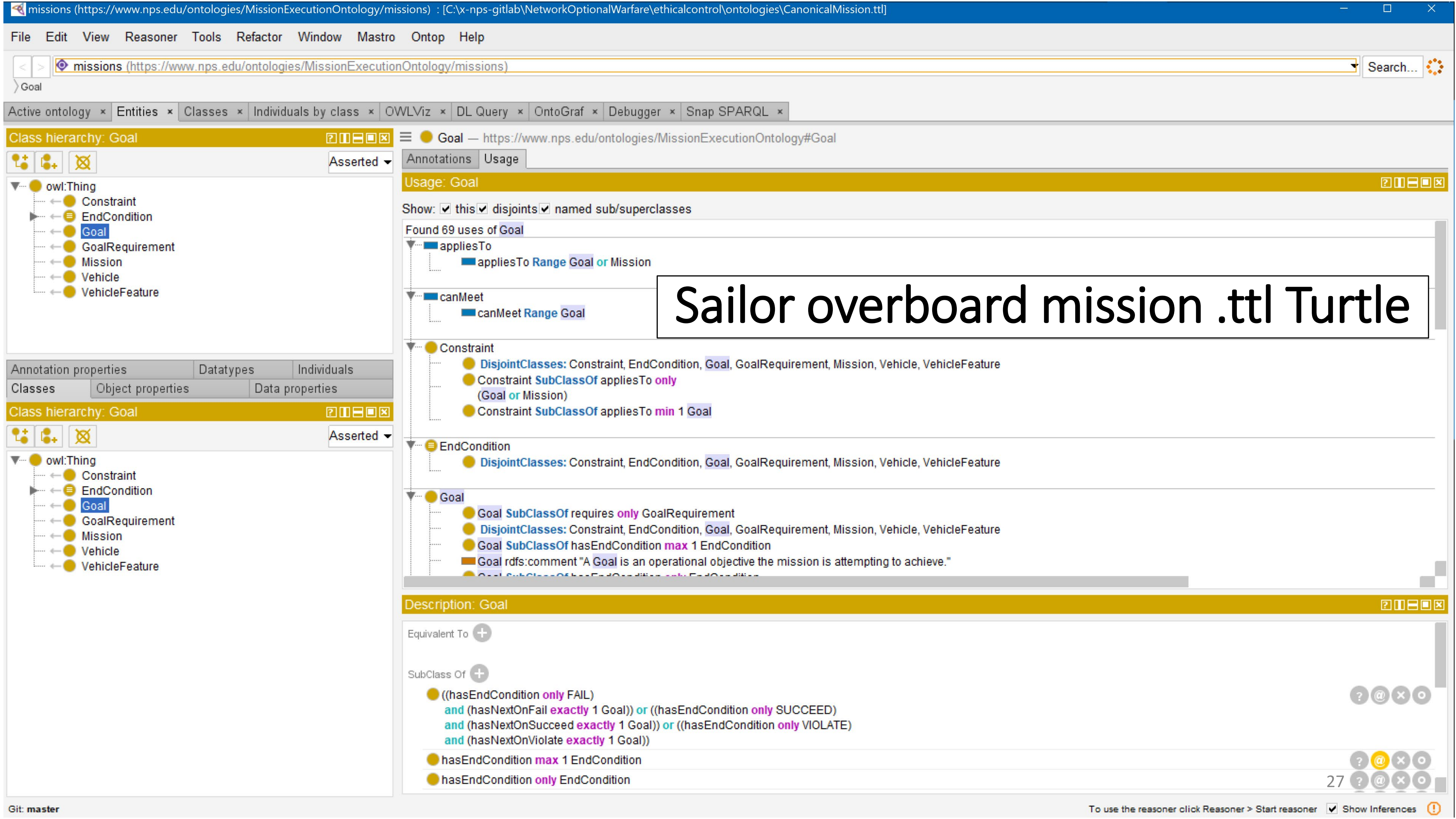
SailorOverboard.xml 5.63 KB

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





Sailor overboard mission .ttl Turtle
missions (https://www.nps.edu/ontologies/MissionExecutionOntology/missions) : [C:\x-nps-gitlab\NetworkOptionalWarfare\ethicalcontrol\ontologies\CanonicalMission.ttl]
File Edit View Reasoner Tools Refactor Window Mastro Ontop Help
missions (https://www.nps.edu/ontologies/MissionExecutionOntology/missions)
Search...
Goal
Active ontology Entities Classes Individuals by class OWLViz DL Query OntoGraf Debugger Snap SPARQL
Class hierarchy: Goal
Asserted
owl:Thing
Constraint
EndCondition
Goal
GoalRequirement
Mission
Vehicle
VehicleFeature
Annotation properties Datatypes Individuals
Classes Object properties Data properties
Goal — https://www.nps.edu/ontologies/MissionExecutionOntology#Goal
Annotations Usage
Usage: Goal
Show: this disjoints named sub/superclasses
Found 69 uses of Goal
appliesTo
appliesTo Range Goal or Mission
canMeet
canMeet Range Goal
Constraint
DisjointClasses: Constraint, EndCondition, Goal, GoalRequirement, Mission, Vehicle, VehicleFeature
Constraint SubClassOf appliesTo only (Goal or Mission)
Constraint SubClassOf appliesTo min 1 Goal
EndCondition
DisjointClasses: Constraint, EndCondition, Goal, GoalRequirement, Mission, Vehicle, VehicleFeature
Goal
Goal SubClassOf requires only GoalRequirement
DisjointClasses: Constraint, EndCondition, Goal, GoalRequirement, Mission, Vehicle, VehicleFeature
Goal SubClassOf hasEndCondition max 1 EndCondition
Goal rdfs:comment "A Goal is an operational objective the mission is attempting to achieve."
Description: Goal
Equivalent To
SubClass Of
((hasEndCondition only FAIL)
and (hasNextOnFail exactly 1 Goal)) or ((hasEndCondition only SUCCEED)
and (hasNextOnSucceed exactly 1 Goal)) or ((hasEndCondition only VIOLATE)
and (hasNextOnViolate exactly 1 Goal))
hasEndCondition max 1 EndCondition
hasEndCondition only EndCondition
Git: master
To use the reasoner click Reasoner > Start reasoner
Show Inferences

https://gitlab.nps.edu/Savage/EthicalControl/blob/master/missions/turtle/SailorOverboardConverted.ttl

More

SailorOverboardConverted.ttl 2.3 KB

```
### Namespace declarations
@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> . # TODO consider URI matching URL

### Import base ontology
<https://www.nps.edu/ontologies/MissionExecutionOntology/missions> rdf:type owl:Ontology ;
owl:imports <https://www.nps.edu/ontologies/MissionExecutionOntology> .

#################################################################
#    Individuals
#################################################################

:SailorOverboard rdf:type owl:NamedIndividual , owl:Thing ;
  meo:startsWith :Goal1 ;
  rdfs:comment "AVCL mission to perform SAILOR OVERBOARD operations, carried out in concert with shipboard emergency procedures." .

:Goal1 rdf:type owl:NamedIndividual , owl:Thing ;
  meo:hasNextOnFail :Goal2 ;
  meo:hasNextOnSucceed :Goal2 ;
  meo:hasNextOnViolate :Goal5 ;
  rdfs:comment "Deploy, Launch" .

:Goal2 rdf:type owl:NamedIndividual , owl:Thing ;
  meo:hasNextOnFail :Goal3 ;
  meo:hasNextOnSucceed :Goal4 ;
  meo:hasNextOnViolate :Goal5 ;
  rdfs:comment "Rendezvous with Sailor" .

:Goal3 rdf:type owl:NamedIndividual , owl:Thing ;
  meo:hasNextOnFail :Goal5 ;
  meo:hasNextOnSucceed :Goal4 ;
  meo:hasNextOnViolate :Goal5 ;
  rdfs:comment "Search for Sailor " .
```

[SailorOverboardConverted.ttl](https://gitlab.nps.edu/Savage/EthicalControl/blob/master/missions/turtle/SailorOverboardConverted.ttl)

Turtle excerpt autogenerated in gitlab.nps.edu version control

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"

TODO: SHACL queries

Enablers for broad outcomes

Military perspective is subset of broader shared global challenges

Imperatives are fundamentally important, thus any solutions must scale up

Semantic Web: Knowledge Representation at Web Scale

- Architects of the World Wide Web have laid out a layered set of standards to achieve the Semantic Web vision: “not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation” (Berners-Lee et al. 2001)
- Ultimate goal: achieve a ***scalable trusted information infrastructure*** where humans and software interact meaningfully, in a repeatable environment where expectations of quality and integrity are met.
- Scalable approach indicates that single (ship + robot) solutions have potential to grow and encompass many simultaneous systems, thus improved data sharing, mission deconfliction, coordinated operations

Semantic Web Stack

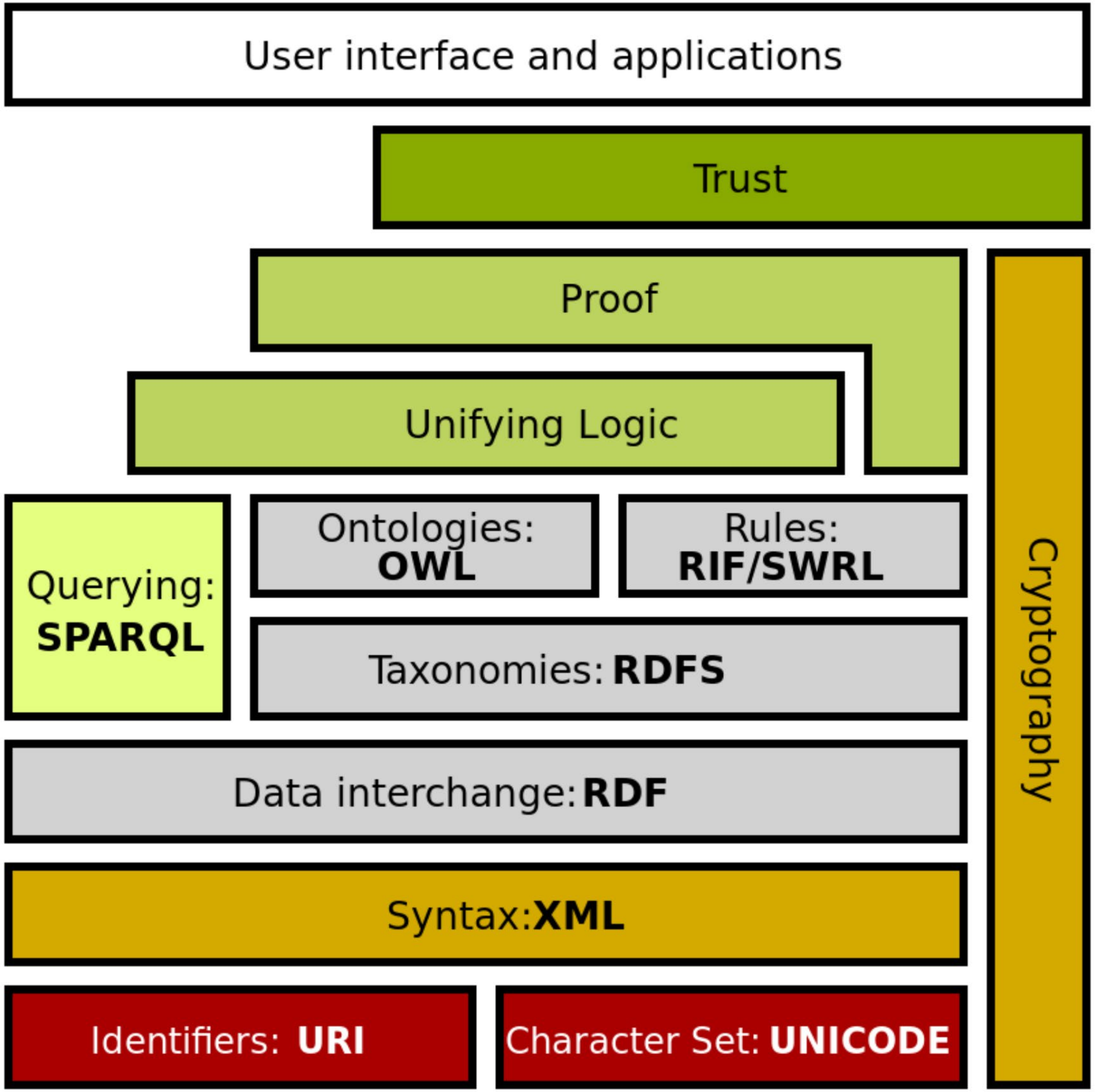
Extends larger Web architecture

- 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.

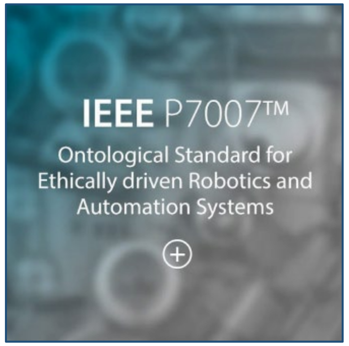


IEEE P7000-series Standards Projects

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



Published November 2021!

- 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.

How can we measure all that?

Metrics, and testing methodology, are essential and perhaps sharable

Where Navy needs to go next…

- Massive testing of unmanned hardware + software ability to follow both orders and constraints in physically realistic virtual environments: **TestDevOps** for learning-improvement cycles.
- **Influence and impact operations**: certify real-world capabilities via field experimentation (FX), confirmed by Big Data analysis, defense test range exercises, wargames and theses.
- Human warfighters and mission commanders (not just engineers) review and approve unmanned systems as… **qualified.**
- New normal will be **human + machine teaming** together; future must mainstream all capabilities in acquisition and deployment.
- **Welcome to the future** – the horizon is here!

Workshop day 3 lookahead: discussion invitation

We will discuss Meaningful Metrics for Demonstrating Ethical Supervision of Unmanned Systems, showing design framework based on

- "How do you measure that?" as critical question, our problem domain requires mission-oriented testing repeatably and indefinitely
- Live-virtual-constructive (LVC) TestDevOps for spiral improvement

We are listening closely and studying each of your contributions… thanks!

Your inputs are welcome on these cross-cutting workshop themes.

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.
- ***How?*** Testing testing testing: simulation, virtual environments, analytics, field experiments, real-world analytics, new normal...

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
https://nps.edu/faculty-profiles/-/cv/clblais

Code MV/Bl, Naval Postgraduate School

Monterey California 93943-5000 USA

1.831.656.3215 work