updated 3 August 2022

Integrating Dimensions of Autonomous Decision Making (DADM) Risk Elements into Unmanned System Mission Orders

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

1. **Overview**. This paper describes how terms of reference and naming conventions were applied to formally define Dimensions of Autonomous Decision Making (DADM) concepts and risk elements as part of Autonomous Vehicle Command Language (AVCL) and Mission Execution Ontology (MEO). These integrated capabilities allow identification of risks and mitigations for mission plans and orders for unmanned systems.

2. **Motivation.** The report Dimensions of Autonomous Decision-making (DADM) by Center for Naval Analyses (CNA) has introduced a process designed to identify and potentially mitigate risks from the military use of artificial intelligence. In effect, it also presents a taxonomy of concepts that can be used for semantic characterization of goals and constraints within structured mission orders for unmanned systems. Figure 1 shows excerpts.

DAD#1: Standard semantics and concepts

- Have all parties identified all the important terms being used in the development and use of the IAS that require definition?
- Are all parties (when they come from different organizations with different doctrine with respect to IAS use) using consistent and non-conflicting doctrinal terminology?
- Does IAS use require the use of rapidly emerging terminology that must be defined and agreed upon before use?
- Are all parties using the same definitions for "artificial intelligence," "intelligent autonomous systems," "autonomy," "automatic," and "autonomous functionality?"
- Are all parties using the same definitions for "peacetime status" and wartime status?"
- Are all parties using the same definitions for IAS "degree of autonomy?"
- Are all parties using the same definition for "realistic operational environment" for IAS developmental and operational test and evaluation purposes?
- Are all parties using the same definitions for "training data, input data and feedback data?"
- Are all parties using the same definitions for the several and distinct operational phases?[27]
- Are all parties using the same risk management framework?
- Are all parties using the same technical standards throughout the entire lifecycle of the IAS?
- Are all parties using the same metrics for quantitative analysis (e.g., analyzing confidence levels, comparing similarities, measuring differences)?

DAD#2: Continuity of legal accountability

Pre-operational phase considerations

- Has the concept of operation been analyzed to verify that no temporal or spatial accountability gaps exist regarding the use of IAS?
- Has the concept of operation been analyzed to verify that no transfer of command and control over the IAS can occur without specific authorization by the person(s) designated to be accountable for the use of the IAS?
- Is/are the person(s) designated to be accountable for the use of an IAS the only person(s) with the physical ability to transfer decision-making capability to the IAS?
- Is/are the person(s) designated to be accountable for the use of an IAS the only person(s) able to authorize in situ changes to the IAS's configuration created by exposure to incoming data streams?
- Have all persons who may be designated to be accountable for the use of an IAS received training on the Law of Armed Conflict and ethics policies?
- Have all persons who may be designated to be accountable for the use of an IAS been briefed on the current and prevailing rules of engagement?
- Have all persons who may be designated to be accountable for the use of an IAS understand that transfer of decision-making capabilities to an IAS does not transfer accountability for the results of any decisions made by that IAS?
- Can IAS lethal capabilities (or other capabilities authorized only for use during wartime) be disabled during peacetime and only be activated after a verifiable transmission is received from an accountable (military or civilian) authority?
- Can the IAS engagement parameters be pre-set to either allow or prohibit it from developing its own target selection, discrimination, or engagement criteria?

Figure 1. Dimensions of Autonomous Decision-making (DADM) is structured with 13 top-level concept categories containing corresponding conceptual subcategories and numerous risk elements.

Motivated by Ethical Control of Unmanned Systems, the preliminary work reported here has explored precise naming and representation of DADM concepts as Semantic Web relationships. This is a necessary step for usage within Autonomous Vehicle Command Language (AVCL) mission definitions and corresponding Mission Execution Ontology (MEO) semantic validation of taxonomy preconditions, postconditions, and traceable adherence within the validatable conduct of diverse unmanned-system missions. The DADM risk elements appear to be an excellent fit.

This note describes representations for DADM risk elements using Semantic Web constructs. Continued work is applying specific risk elements to common mission goals performed by unmanned systems, permitting corresponding risk-aware directions and due-diligence oversight by human watchstanders.

3. **Multiple simultaneous design goals.** Based on an extremely large set of reported real-world dependencies, the DAD offers a thorough set of concepts for consideration. Several steps are needed in order to elevate this work into a formal taxonomy suitable for Semantic Web use. Design goals are primarily oriented around clear naming of each risk element in order to achieve both human and machine readability. Stanford University's popular [Protégé application]() provides a thorough Semantic Web authoring environment to explore and demonstrate all of these characteristics. Since the published risk elements do not include specific names, we are focusing closely on how identifiers might be defined for best effect.

Current naming conventions reflect initial representational experimentation on our part.

- *Uniqueness* of names to enable unambiguous semantic references,
- *Hierarchical* class/subclass structure to match CNA DADM reference document taxonomy,
- *Descriptive clarity* for human comprehension of CamelCase terms and names,
- *Brevity* of rdfs:label identifiers for corresponding rule-definition rigor and clarity, as well as future extensibility with multilingual labels to permit strict and coherent usage by international partners,
- *Embedded references* that correspond exactly to structure of authoritative CNA document, supporting both end-user comprehension and full traceability of original requirements during mission analysis.

The [Protégé application]() provides excellent immediate feedback on whether such rule structures meet validity requirements. A validation stylesheet provides further consistency checking of the ontology for completeness and consistency. Future work includes exploration of Shapes Constraint Language (SHACL) for even-deeper validation of semantic relationships.

4. Ontology structural design and Semantic Web results

Work constructing this ontology proceeded with four Ph.D. students studying Semantic Web, proceeding together to refine naming and design rules for clarity and consistency. Mostly similar to (but sometimes different from) object-oriented model design, incremental development and refinement was essential. This was a good activity further exercising our ability to express concepts clearly and succinctly in accordance with Web Ontology Language (OWL) and related standards by World Wide Web Consortium (W3C).

Numerous concerns pertain for declarative mission design and eventual composable Semantic Web query. Our primary guiding reference:

- Dean Allemang, James Hendler, and Fabien Gandon, *Semantic Web for the Working Ontologist: Effective Modeling for Linked Data, RDFS, and OWL*, ACM Books, third edition, 2020. https://dl.acm.org/doi/book/10.1145/3382097

The following excerpts show naming-convention examples using DADM categories and risk elements.

- Define unique ontology classes for **565 consolidated risk elements**, consolidated and distilled from **4641 diverse risk elements** originally reported by a large number of real-world programs**.**
- Class names are terse, while rdfs:label entries are more verbose. This matches OWL best practices for internationalization (I18N) of explanatory rdfs:label values, permitting consistent syntactical/semantic validation by all Semantic Web tools with corresponding tooltip hints in the appropriate language for human end users (English, French, German, Polish, Ukrainian, Korean, Japanese, and so forth). Each rdfs:label starts with same name to avoid ambiguity.
- Example: class **DADM01** has rdfs:label **DADM01_StandardSemanticsConcepts**

- Prefix taxonomy terms: **DADM02_ContinuityLegalAccountability**
- Prefix subcategory terms: **DADM02.1_PreOperationalPhaseConsiderations**
- Prefix risk elements: **RE02.1a_SpatialTemporalAccountabilityGaps**
- Underscore character _ is separator between referential prefix and term name, providing unique readable long-form identifiers,
- Prefix term saved as *rdfs:label* as tooltip, e.g. **DADM02**, **M02.1**, **RE02.1a** etc.
- Prose definition as *rdfs:comment* provides full verbatim description, linked within semantic class definition,
- Unique document reference identifier **DRM-2021-U-03064201Rev** as *owl:versionInfo* in topmost **DimensionsAutonomousDecisionMaking** class, providing global context,
- Document page number included as *owl:versionInfo* for each term and element, supporting contextual human lookups,
- Note that CamelCaseName synopses and corresponding alphanumeric identification were defined during manual construction of these OWL classes from the CNA documentation. Future DADM document definitions might benefit from such clearer naming and numbering.
- Also adding sections for both Abbreviations and References so that cross-referencing documentation and queriability might further enrich the value of this knowledge base.
- These CamelCase synopsis names are initial "best efforts" and can benefit from broader scrutiny by authors and readers of original CNA report.

.This work is also suitable for precise semantic correlation to future versions of [Dimensions of Autonomous Decision-making (DADM)](). Figure 2 shows initial excerpts from our example implementation, with addition of further risk elements, abbreviations and cross-references proceeding carefully.

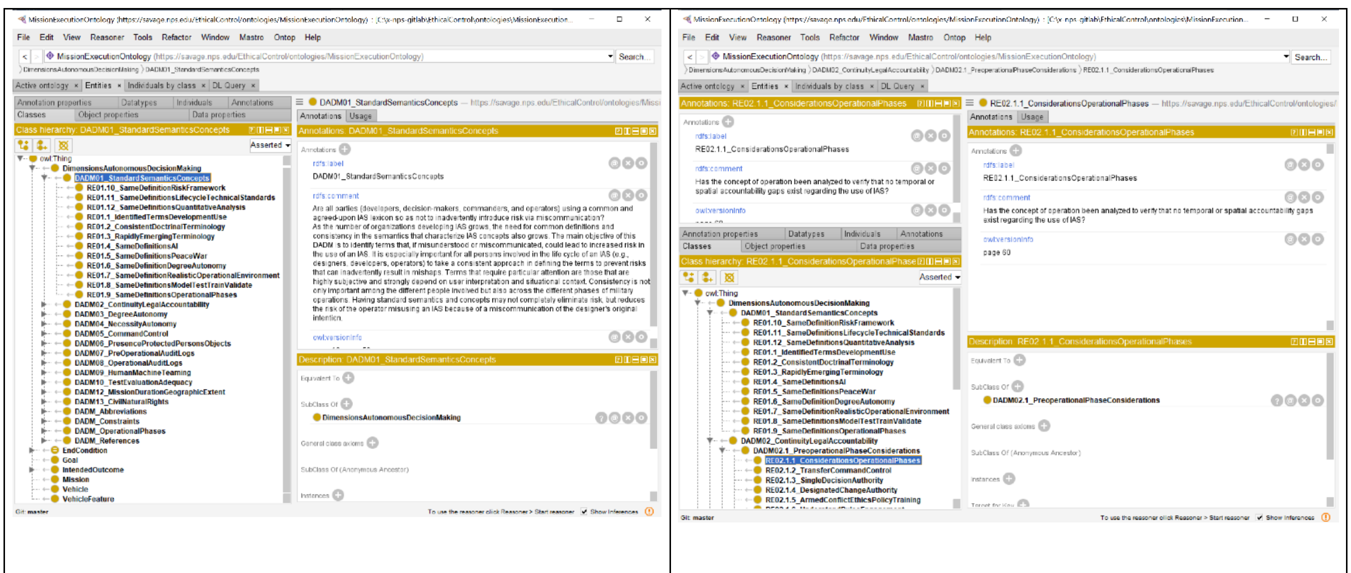

Figure 2. Providing unique names for DADM risk elements is an essential step, unlocking potential for unambiguous referencing both by humans and Semantic Web rule bases. These screenshots illustrate creation of a DADM ontology using the Stanford [Protégé]() application, which are then correlated with Autonomous Vehicle Command Language (AVCL) Mission Execution Ontology (MEO).

We expect that prefixes, names and corresponding descriptions may change slightly over time as practice and usage provides further insight. One guiding heuristic is ironic but also measurable and effective:

- "You know you have the right name when… no one asks about it anymore."

5. Ontology Production Details

Multiple processes were used to produce DADM ontology for integration into the Mission Execution Ontology (MEO), and then produce corresponding enumerations for the AVCL schema. Although our team usually first produces data models before ontologies, in this case the originating DADM reference was most clearly suited for ontology production.

The team produced Semantic Web constructs using Protégé, which does an excellent job maintaining complex internal relationships. Initial rules were saved in gitlab version control, allowing careful scrutiny of changes as the design evolved. Version control then permitted a divide-and-conquer approach for each student to work on independent DADM categories, with careful integration then possible. Saving the intermediate Turtle ontology as an equivalent OWL XML file enabled creation of an XSLT stylesheet for automated checking of consistent naming patterns. Excerpts follow.

Terse Triple Language (Turtle) excerpt

```
###  https://savage.nps.edu/EthicalControl/ontologies/MissionExecutionOntology#RE02.1.1
meo:RE02.1.1 rdf:type owl:Class ;
             rdfs:subClassOf meo:DADM02.1 ;
             rdfs:comment "Has the concept of operation been analyzed to verify that no
temporal or spatial accountability gaps exist regarding the use of IAS?" ;
             rdfs:label "RE02.1.1_ConsiderationsOperationalPhases" ;
             owl:versionInfo "page 60" .
```

Web Ontology Language (OWL) XML excerpts

```
    <Declaration>
        <Class IRI="#RE02.1.1"/>
    </Declaration>
    <SubClassOf>
        <Class IRI="#RE02.1.1"/>
        <Class IRI="#DADM02.1"/>
    </SubClassOf>
    <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:comment"/>
        <IRI>#RE02.1.1</IRI>
        <Literal>Has the concept of operation been analyzed to verify that no temporal or
spatial accountability gaps exist regarding the use of IAS?</Literal>
    </AnnotationAssertion>
    <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#RE02.1.1</IRI>
        <Literal>RE02.1.1_ConsiderationsOperationalPhases</Literal>
    </AnnotationAssertion>
    <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="owl:versionInfo"/>
        <IRI>#RE02.1.1</IRI>
        <Literal>page 60</Literal>
    </AnnotationAssertion>
```

XSLT stylesheet for naming-pattern confirmation excerpt

```
        <!-- autogenerated dadmRiskElementType enumerations -->
        <xsd:simpleType name="dadmRiskElementNames">
                <xsd:annotation>
                        <xsd:documentation>DADM Risk Elements</xsd:documentation>
                </xsd:annotation>
                <xsd:restriction base="xsd:string">
                        <xsd:enumeration value="RE02.1.1_ConsiderationsOperationalPhases">
                                <xsd:annotation>
                                        <xsd:appinfo>Has the concept of operation been analyzed
to verify that no temporal or spatial accountability gaps exist regarding the use of
IAS?</xsd:appinfo>
                                </xsd:annotation>
                        </xsd:enumeration>
                </xsd:restriction>
                <!-- additional enumerations -->
        </xsd:simpleType>
```

Output checker excerpt shows no warnings/errors:

```
MeoDadmNamingVerification ontologies/MissionExecutionOntology3.1-DADM.owl using
stylesheets/MeoDadmNamingVerification.xslt
MeoDadmNamingVerification.xslt checks for MEO DADM .owl ontology:

Class IRI='#RE02.1.1'
   owl:subClassOf    #DADM02.1
   rdfs:label        RE02.1.1_ConsiderationsOperationalPhases
   owl:versionInfo   page 60
   rdfs:comment     "Has the concept of operation been analyzed to verify that no temporal
or spatial accountability gaps exist regarding the use of IAS?"
```

***Figure 3**. Text processing excerpts showing (a) Turtle rules manually produced using Protege, (b) corresponding OWL RDF excerpt, (c) XSLT stylesheet for naming-pattern confirmation, and (d) example stylesheet-verification output.*

Once about 40% of the DADM rules and risk elements were established satisfactorily and naming-patterns confirmed, the team performed the generation of plain-text Turtle triples using a Python script and iterative processing. Common techniques for "big data" processing were applied. The process began with cutting and pasting relevant prose from the DADM pdf source document text into a text file. Specifically, annotations are made to the source text so a Python script can generate the nodes with the appropriate relationships. These include adding variable names before the Risks (e.g., *IdentifiedTermsDevelopmentUse* on line 3), adding additional bullets to indicate hierarchy (e.g., * addition on line 18), and placing @ symbols before text line parents that are not risk elements (e.g., @on line 17). Finally, text cleanup including removal extra new line characters and put page numbers at the beginning vs. the end of the page (e.g., as done on lines 1 and 15).

```
1   PAGE_59
2   #01: Standard semantics concepts
3   • IdentifiedTermsDevelopmentUse|Have all parties identified all the important terms being used in
4   • ConsistentDoctrinalTerminology|Are all parties (when they come from different organizations wit
5   • RapidlyEmergingTerminology|Does IAS use require the use of rapidly emerging terminology that mu
6   • SameDefinitionsAI|Are all parties using the same definitions for "artificial intelligence," "in
7   • SameDefinitionsPeaceWar|Are all parties using the same definitions for "peacetime status" and w
8   • SameDefinitionDegreeAutonomy|Are all parties using the same definitions for IAS "degree of auto
9   • SameDefinitionRealisticOperationalEnvironment|Are all parties using the same definition for "re
10  • SameDefinitionsModelTestTrainValidate|Are all parties using the same definitions for "training
11  • SameDefinitionsOperationalPhases|Are all parties using the same definitions for the several and
12  • SameDefinitionRiskFramework|Are all parties using the same risk management framework?
13  • SameDefinitionsLifecycleTechnicalStandards|Are all parties using the same technical standards t
14  • SameDefinitionsQuantitativeAnalysis|Are all parties using the same metrics for quantitative ana
15  PAGE_60
16  #02: Continuity of legal accountability
17  • @PreoperationalPhaseConsiderations|Pre-operational phase considerations
18  •• ConsiderationsOperationalPhases|Has the concept of operation been analyzed to verify that no t
19  •• TransferCommandControl|Has the concept of operation been analyzed to verify that no transfer o
    accountable for the use of the IAS?
```

Figure 4. Text extraction and preparation for Turtle triple production (file risk_elements_input.txt).

The Python script *generate_turtle.py* iterates through each line and updates variables needed for the reference page number and heading or generates and outputs a line of turtle format text. Interestingly this part of the processing was possible using a Web browser and the Jupyter environment for Python programming.

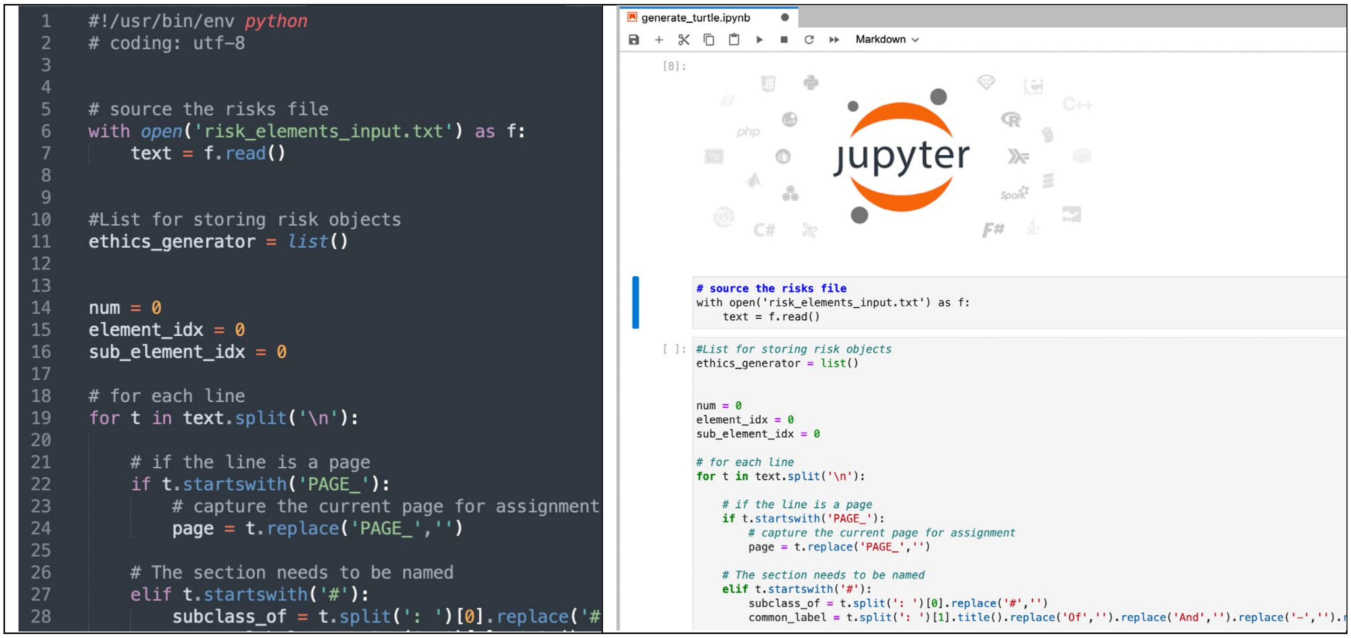

Figure 5*.* *Python programming with Jupyter Web-browser development environment was used to convert original DADM PDF-document prose into Turtle triples, following naming patterns established during initial manual ontology design.*

6. Integrating DADM Risk Elements in AVCL Mission Definitions

The next step performed was integrating these constructs within AVCL mission declarations and MEO ontology (specifically the Constraints hierarchy) for ease of use in referring to individual DADM categories and corresponding Risk Elements to facilitate further testing. Careful prior ontology design was indeed shown to be adaptable for eventual separation into a standalone set of relationships. Excerpts from current results are shown in Figure 5.

Production of XML enumerations for the AVCL XSD schema performed an unusual round trip: going from ontology back to XML schema. Once again a stylesheet was used to re-express ontology constructs, this time going from OWL XML to XSD XML. Results were manually copied into the AVCL 3.1 schema, tested and then added to version control for long-term reuse.

Current work includes applying the semantics of DADM Risk Elements as expected risks in corresponding mission goals. This is expected to be a lengthy process as numerous missions, contexts, goals and risks are considered. In the long term, it is possible that creation of mission libraries and templates might facilitate the comprehensive and orderly consideration of risks as part of operational planning for robot missions.

Achieving effective human supervision of remote, autonomous robot operations is widely considered among the most significant challenges facing the ethical control of unmanned systems. As such practices are considered and tested using AVCL, MEO and DADM together, it is important to note that the chosen representations using Semantic Web standards hold the technical capability for arbitrarily large scalability. Thus mission validation and testing appears feasible both for individual ships and broader fleet aggregations simultaneously.

XSD XML Schema excerpt

```
	<!-- autogenerated dadmRiskElementType enumerations -->
	<xsd:simpleType name="dadmRiskElementNames">
		<xsd:annotation>
			<xsd:documentation>DADM Risk Elements</xsd:documentation>
		</xsd:annotation>
		<xsd:restriction base="xsd:string">
			<xsd:enumeration value="RE02.1.1_ConsiderationsOperationalPhases">
				<xsd:annotation>
					<xsd:appinfo>Has the concept of operation been analyzed
to verify that no temporal or spatial accountability gaps exist regarding the use of
IAS?</xsd:appinfo>
				</xsd:annotation>
			</xsd:enumeration>
		</xsd:restriction>
		<!-- additional enumerations -->
	</xsd:simpleType>
```

AVCL XML Excerpt: Search Goal with Risk Element

```
<Goal id="Goal2" title="Perform racetrack in Delmar Boat Basin"
      description="Provide acoustic signals to MEMS sensor" phase="Patrol"
      nextOnSuccess="Goal3" nextOnFailure="Goal3" nextOnException="Goal3">
	<Search datumType="area" requiredPD="0.5">
		<Target name="Orbit using long rectangular track" id="Racetrack"/>
	</Search>
	<Risk riskName="RE11.4.1_UnderstandingReasoningMethodsLimitations"
		mitigation="watchstander training" mandatory="true"/>
	<OperatingArea refid="NpsMemsOperatingArea"/>
	<Duration value="6000"/>
</Goal>
```

***Figure 6**. XSD XML Schema excerpt showing validation of DADM risks as enumeration datatypes, and (b) example AVCL mission excerpt illustrating example usage in an unmanned system mission. Presence of a Risk is required to have a noted mitigation by human operators.*

7. Example Mission Definitions for Robot Risk Assessment and Mitigation by Human Supervisors

Expected next steps include an ongoing series of graduate-student exercises, focused on well-structured mission authoring and semantic exploration of DADM concepts within plausible AVCL/MEO real-world scenarios of interest.

Current academic work in NPS course IS3460 Networked Autonomous Unmanned Systems is exploring definition of a variety of robot missions. This mission-modeling work in progress will test the sensible inclusion of DADM Risk Elements as part of risk-mitigation support to assist officers responsible for defining missions with potential for lethal (or life-saving) force.

Figure 6 shows a simple field experiment for sensor monitoring of an unmanned-system tracks. The three-way ternary logic (success-failure-exception) allows mission planners to consider all possible decision branches with in mission. The mission-flow goal diagram is followed by Figure 7 which shows the same mission formally expressed as a human-readable, machine-readable, validated AVCL XML.

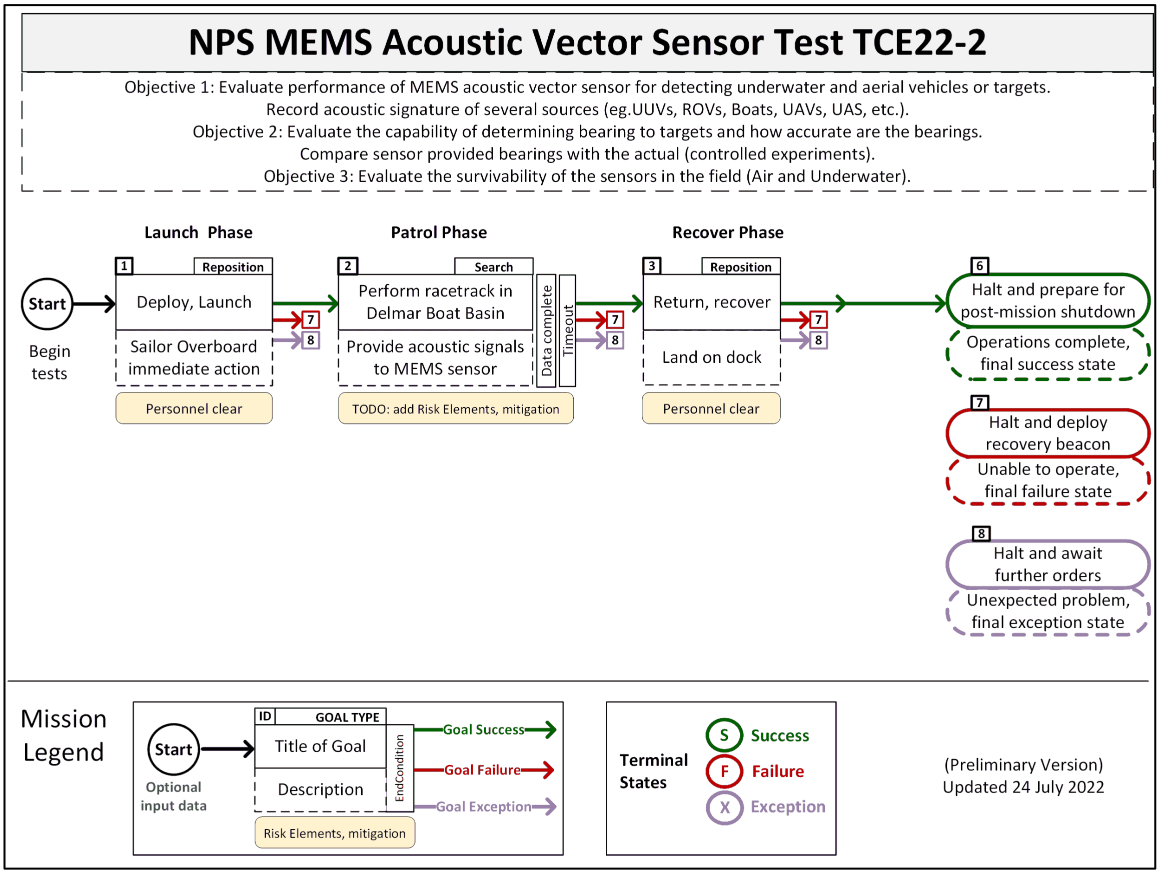

Figure 6. Preliminary mission design showing assignment of Risk Elements and mitigations to AVCL mission goals.

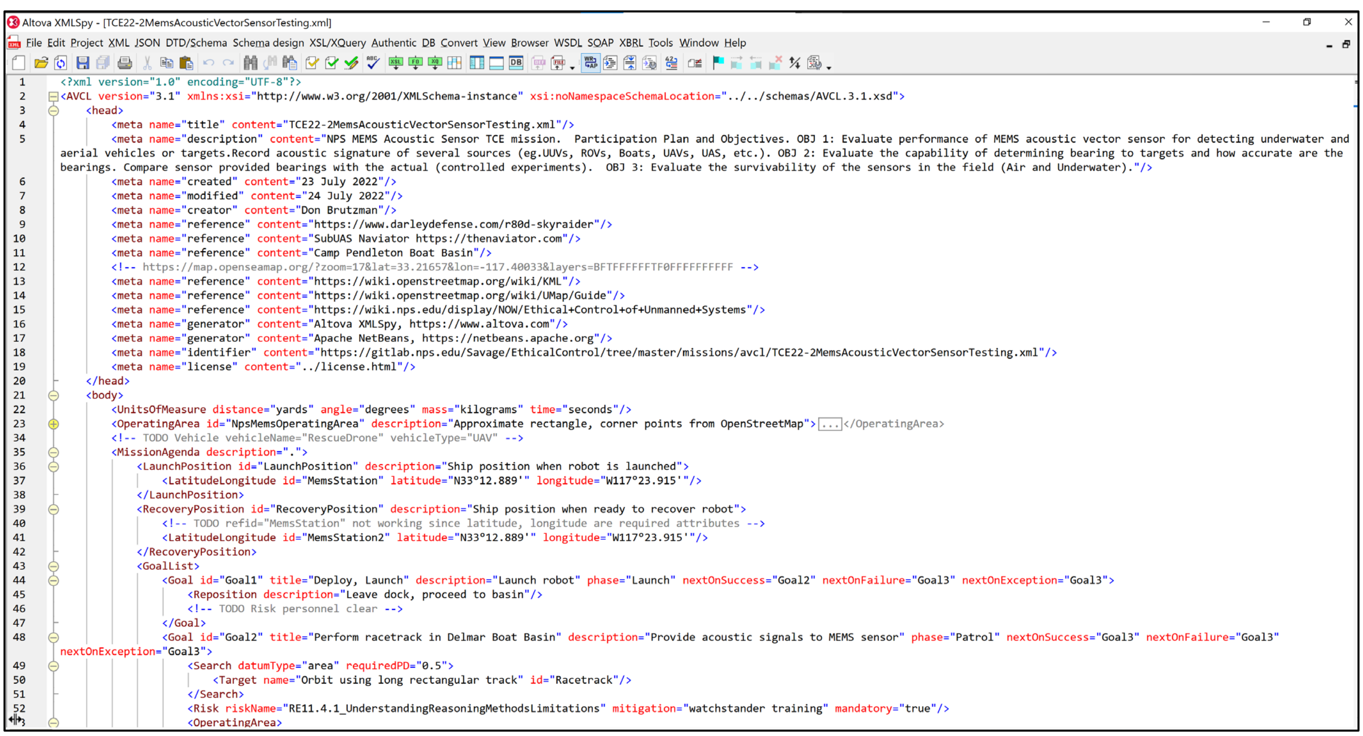

Figure 7. Example AVCL mission showing unmanned-system goals annotated with DADM Risk Elements (REs).

Autonomous Vehicle Command Language (AVCL) terms of reference for precise description of unmanned system goals and tasks continue to be refined through practice. Conceptual testing continues by NPS graduate students examining typical robot missions performed in diverse venues for field experimentation. This work is becoming part of recurring curriculum efforts. Course synopsis and selected student-project goals follow in Figure 8.

IS3460 Networked Autonomous Vehicles and Unmanned Systems (4-0)

July-September 2022

Synopsis. This unclassified course examines autonomous and unmanned systems and platforms from a systems and operational perspective. Historical and modern systems are discussed to include Industrial Control Systems, botnets, UxVs, etc. The nature of autonomy versus unmanned systems is examined. Opportunities and security issues presented by the growing dependence upon these systems and platforms are studied. The ethics of using unmanned and autonomous platforms and systems for warfare is examined, along with the ethics of attacking such systems when integrated into society. Operational applications within the private and public sectors, as well as the military, are discussed. Current and future research into autonomy is examined. Prerequisites: none.

Define an **unmanned system mission**. Report on it, working pairwise on partnered missions is good.

- Describe the mission and unmanned-system capabilities needed to perform it.
- Why do you think it is important?
- List the input contraints, goal outcomes and expected goal steps within your mission.
- What networking is required? What degree of independent machine autonomy is needed?
- Try to describe the mission using Autonomous Vehicle Control Language (AVCL).
- How can human operators maintain positive control over your unmanned system's mission?
- Are there appropriate Risk Elements in the Dimensions of Autonomous Decision Making (DADM) Ontology that can be aligned as warrior checkpoints when approving robot goals?

Directly observe a tested system in one of the **field experiments** (in person or virtually). Report on it.

- What were their goals? How did it go? Did it work? Is it repeatable? What is next?
- Over and above: mission storyboard in Microsoft project and in SPIDERS3D virtual environment.

Figure 8. Course synopsis for IS3460 Networked Autonomous Vehicles and Unmanned Systems, including applied course-project tasks.

8. **Conclusions.** The DADM taxonomy is representable as a formal ontology, completely and unambiguously, in a manner that is both human readable and machine readable. Precise testing is necessary for exercising, evaluating, and improving this set of properties.

9. **Recommendations.** continued work is warranted to demonstrate the taxonomy of DADM risk elements as a Semantic Web ontology. Formal constraints might then show risk-element relationships within unmanned-system mission definitions using AVCL and MEO for Ethical Control of Unmanned Systems. Defining and testing a steadily growing set of example missions will steadily lead to further clarity and usefulness.

Acknowledgements. DADM ontology contributions by Garrett Cookson, Michael Holihan, Jessica Kimball, and Rick Lentz are gratefully acknowledged.