Universal Core Semantic Layer

Barry Smith, Lowell Vizenor and James Schoening
National Center for Ontological Research, University at Buffalo
701 Ellicott Street, Buffalo NY, 14214
Army Net-Centric Data Strategy Center of Excellence, Building 1209, Fort Monmouth, NJ 07703
phismith@buffalo.edu, lowell.vizenor@ctg.com, james.schoening@us.army.mil

***Abstract* – The Universal Core (UCore) is a central element of the National Information Sharing Strategy that is supported by multiple U.S. Federal Government Departments, by the intelligence community, and by a number of other national and international institutions. The goal of the UCore initiative is to foster information sharing by means of an XML schema providing consensus representations for four groups of universally understood terms under the headings *who*, *what*, *when*, and *where*. We here describe a project to create an ontology-based supporting layer for UCore, entitled 'Universal Core Semantic Layer' (UCore SL), and describe how UCore SL can be applied to further UCore's information sharing goals.**

***Index Terms* – Ontology, Data Integration, Semantic Technology, OWL DL, Universal Core**

I. The Universal Core

The Universal Core (UCore) [1] is a US Federal Government information sharing initiative that is supported by the US Departments of Defense, Energy, Justice, and Homeland Security, by the Intelligence Community, and by a large number of other national and international agencies. UCore supports the principles of the Department of Defense (DoD) and Intelligence Community (IC) Data Strategies by defining a small set of common data elements that are implemented in a lightweight information exchange schema that is shared across multiple agencies.

The prime focus of the UCore initiative is messaging. UCore is designed to promote information sharing across multiple message domains by means of a simple XML message format built on a taxonomical structure comprising four groups of terms under the headings *who*, *what*, *when*, and *where*. Table 1, below, represents the taxonomy as released in UCore Version 2.0, which is the version upon which we focus in what follows. Table 2 represents the relations contained within the UCore 2.0 xsd:schema.

The UCore strategy is to require message-creators to construct for each message a *digest*, a summary built out of a restricted vocabulary of UCore terms, and to link elements from the message payload to this digest. Developers of information systems are encouraged to use these terms wherever practical in order to realize the goal of facilitating automated sharing of information within and across agencies. To reap maximal benefit from its messaging resources, participants in the UCore initiative offer validation processes and tools intended to promote machine understanding of message content, thereby enabling multiple different types of information retrieval, reasoning and consistency checking.

The UCore taxonomy consists of terms (such as 'Person' or 'Organization') which are universally understood in the sense that they require no domain-specific expertise for their understanding. The taxonomy can thereby be shared by many different types of users, and thus it provides the opportunity for interoperability over many different sorts of domain-specific exchanges. As M. Daconta expresses it:

> if I have a UCore-wrapped National Information Exchange Model [NIEM] message from Immigration and Customs Enforcement about illegal immigrants wounded during criminal activity and I have a UCore-wrapped Health and Human Service Department message on visitors to emergency rooms, I have enabled immediate cross-domain search. … UCore is a process of extracting cross-domain commonality from your message flows, thereby massively broadening the possible adoption and use of your shared information. In information sharing, adoption by consumers is the key value metric. [2]

The UCore 2.0 taxonomy in its current form is well adapted to realizing this strategy of information sharing on the basis of universally understood terms. UCore 2.0 as a whole, however, still has a number of problems, including a mismatch between this taxonomy and UCore's larger XML schema. The latter includes a number of elements that are not represented in the taxonomy, including spatial and temporal terms:

> GeoLocation: A physical location with coordinates, or a simple geospatial region;
>
> TimeInterval: An interval in time, defined by two instants in time.

Since these elements do not have a corresponding representation in the taxonomy, their intended semantics remain implicit, and no straightforward way exists to link them to, say, spatio-temporal ontologies.

II. UCore and the Army Net-Centric Data Strategy

UCore is designed not only to support messaging and the retrieval and analysis of message content. It is also built in such a way as to support interoperability of information systems of a variety of different types. The strategy is to have UCore serve as the consensus starting point for the construction of successive layers of more inclusive artifacts,

TABLE I: UCORE 2.0 TAXONOMY

uc:Entity		uc:Event	
uc:Cargo	uc:LivingThing	uc:AlertEvent	uc:LawEnforcementEvent
uc:CollectionOfThings	uc:Animal	uc:CommunicationEvent	uc:MigrationEvent
uc:CyberAgent	uc:Person	uc:CriminalEvent	uc:MilitaryEvent
uc:Document	uc:MicroOrganism	uc:CyberSpaceEvent	uc:NaturalEvent
uc:Environment	uc:Plant	uc:DisasterEvent	uc:ObservationEvent
uc:Equipment	uc:Organization	uc:EconomicEvent	uc:PlannedEvent
uc:Facility	uc:PoliticalEntity	uc:EmergencyEvent	uc:PoliticalEvent
uc:FinancialInstrument	uc:Sensor	uc:EnvironmentalEvent	uc:PublicHealthEvent
uc:GeographicFeature	uc:Vehicle	uc:EvacuationEvent	uc:SecurityEvent
uc:GroupOfOrganizations	uc:Aircraft	uc:ExerciseEvent	uc:SocialEvent
uc:GroupOfPersons	uc:GroundVehicle	uc:FinancialEvent	uc:TerroristEvent
uc:InformationSource	uc:Spacecraft	uc:HazardousEvent	uc:TransportationEvent
uc:Infrastructure	uc:Watercraft	uc:HumanitarianAssistanceEvent	uc:WeatherEvent
		uc:InfrastructureEvent	

creating a growing terminology framework within which there can be threaded interoperability corridors tailored to the needs of specialist groups of users.

Against this background, the Army Net-Centric Data Strategy Center of Excellence is supporting experiments to use UCore as the basis for fostering the interoperability of information artifacts created by Communities of Interest (COIs) in the Command and Control (C2) and other domains. The idea is that such COIs will create new vocabularies tailored to meet their unique requirements and thus go beyond the narrow set of UCore terms. By providing an evolving resource of common terms UCore will serve as a central hub designed to maintain a joint community perspective. The long-term goal is that these common terms will create a common reference platform allowing data from diverse COIs to be understood by systems across the DoD and IC. This approach is also designed to allow a level of information sharing between unanticipated users and systems and to reduce the time and cost to implement information sharing across the DoD and IC enterprise, while allowing COIs to focus on their community specific needs.

To achieve these ends, UCore will need to accommodate new requirements from its partner agencies, while at the same time remaining faithful to its key principle of providing a small set of essential terms and relations. This set will however need to be expanded in order to include those universally understandable terms (such as 'weapon') not so far included. UCore has accordingly established a Configuration Control Board (CCB), whose role is to manage change in such a way that successive UCore versions remain useable throughout the change lifecycle.

III. THE OBO FOUNDRY AND BASIC FORMAL ONTOLOGY

The idea of creating consistent extensions on the basis of a common core in order to serve interoperability has been thoroughly explored in the biomedical domain, where the Gene Ontology (GO) was established already in 1998 [3] to provide a resource for the consistent description of biological functions and processes across a multiplicity of different species, including humans. Although initially a logically weakly structured set of terms and definitions, the GO was nonetheless extraordinarily successful in terms of both numbers of users and of the variety of different types of use. As the need began to make itself felt to extend the reach of the GO through the construction of new ontologies designed to serve, for example, the description of clinical phenomena, it was recognized by the GO community that a more systematic approach to logical structure was required in order to ensure cross-domain consistency and thereby enable integration of data across species and disciplines [4].

In 2006, accordingly, the Open Biomedical Ontologies (OBO) Foundry [5] was established, comprising a suite of ontologies built and maintained in such a way as to be interoperable with the GO. Basic Formal Ontology (BFO) [6] plays the role of core for the extension ontologies within this framework [7], so that each ontology is required to employ BFO's restricted set of logically defined ontological relations [8]. Some Foundry ontologies are being created *ab initio* to satisfy the Foundry principles. Legacy ontologies will be subjected to an incremental process of logical reconstruction that designed to ensure that they, and the large quantities of legacy data annotated in their terms, become progressively linked together in a computable way.

IV. UNIVERSAL CORE SEMANTIC LAYER

We describe in what follows an initiative on the part of the Army Net-Centric Data Strategy (ANCDS) [9] Center of Excellence to create an analogous logical infrastructure in support of the UCore endeavor, focusing especially on the application of UCore in the creation of domain and COI-specific extensions. The role of logical core is played in this case by the UCore Semantic Layer (UCore SL), version 1.0 of which was released on June 15, 2009. UCore SL is the product of work by researchers from the National Center for Ontological Research (NCOR) in Buffalo, with considerable input from the intelligence community under the sponsorship of the Office of the Director of National Intelligence (ODNI) CIO.

UCore SL is designed to work behind the scenes in UCore 2.0 application environments as a logical supplement to the UCore messaging standard. Where UCore 2.0 is based on the XML format, in which definitions are logically unarticulated and thus logically based merging of content is not allowed, UCore SL employs the W3C's OWL DL web ontology language, which allows logically articulated definitions to be formulated in such a way as to support such merging. UCore SL offers the entirety of the content UCore 2.0, both taxonomy and relations, in a form which satisfies the needs of users with a need for enhanced logical resources. It provides for logical decomposition of terms and definitions, the ability to reason logically on the basis of the content of these definitions, and thereby also enhanced support for the creation of consistent extension modules. UCore SL is being used as a tool for validation of UCore itself and for the generation of proposals for changes and additions both to UCore 2.0 and to its extensions. It also provides accessibility of UCore message content to W3C-standard OWL-DL technology.

Where UCore 2.0 provides for *syntactic* interoperability through its XML framework and controlled vocabulary, UCore SL offers a logically organized vocabulary of terms, relations and definitions which can serve the *semantic* interoperability of UCore message content.

UCore SL is already helping to provide semantic interoperability in the results of work sponsored by the ANCDS COE on Biometrics and C2 Ontologies carried out by NCOR researchers in Buffalo. We are currently evaluating the ability of UCore SL to provide more powerful reasoning and message-checking capabilities as compared with UCore 2.0 without the added logical support. We and others are also testing the capacities of UCore SL to provide facilities for enhanced data sharing by helping to ensure that extension modules created by different domains or COIs, for example within the C2 framework [10], are created in a logically consistent fashion on the basis of logically sound and easily understood definitions. At the same time C2 and Biometric test extensions are themselves being used to test the adequacy and clarity of UCore SL terms and definitions.

The UCore SL Taxonomy (version 1.0) consists of 144 terms organized into an is-a (subclass) hierarchy, of which, following UCore 2.0, the top two terms are sl:Entity and sl:Event (see table below), corresponding roughly to the continuant and occurrent terms standardly used in upper-level ontologies such as BFO. The UCore SL taxonomy comprehends the entirety of the UCore 2.0 taxonomy in the sense that each one of the 55 terms in the UCore 2.0 taxonomy is mapped to a corresponding UCore SL term. As a result, it is possible to translate UCore 2.0 into UCore SL in order to take advantage of the latter's enhanced logical resources . As UCore itself is expanded, additional resources will be added to UCore SL in order to ensure that this translatability is preserved.

UCore SL contains 16 relations, with definitions relying on those provided in BFO [8]. 12 UCore SL relations have counterparts in UCore 2.0. In keeping with the W3C recommended best practice for reuse of OWL resources, ucore:DistinctFrom and ucore:SameAs are not mapped to corresponding UCore SL relations but rather to owl:differentFrom and owl:sameAs respectively. Four other UCore SL relations taken over from BFO do not correspond to any UCore 2.0 relations but are included in order to ensure logical decomposability of definitions. These are: inheres_in, part_of, participates_in and agent_in.

TABLE II

UCORE 2.0 AND UCORE SL RELATIONS

UCore 2.0 Relations	**UCore SL Relations**
rdfs:subClassOf	rdfs:subClassOf
ucore:AffiliatedWith	slr:affiliated_with
ucore:CauseOf	slr:cause_of
ucore:Controls	slr:controls
ucore:DistinctFrom	owl:differentFrom
ucore:EmployedBy	slr:employed_by
ucore:HasDestinationOf	slr:has_destination_of
ucore:HasFamilialRelationTo	slr:has_familial_relation_to
ucore:HasOriginOf	slr:has_origin_of
ucore:InvolvedIn	slr:involved_in
ucore:LocatedAt	slr:located_at
ucore:OccursAt	slr:occurs_at
ucore:SameAs	owl:sameAs
ucore:SubordinateTo	slr:subordinate_to
ucore:WorksAt	slr:works_at
	slr:agent_in
	slr:inheres_in
	slr:part_of
	slr:participates_in

V. Definitions in UCore 2.0 and UCore SL

The UCore 2.0 definitions are derived primarily from the *Concise Oxford English Dictionary* (OED), which, while helpful to human users, unfortunately only goes part of the way to specifying the intended meaning of the terms in a fashion useful to computers. A further problem with this approach is that there are cases where the provided definition is not in agreement with UCore's own is-a hierarchy. An example is uc:Animal:

> A non-human organism which feeds on organic matter, has specialized sense organs and nervous system, and is able to move about and to respond rapidly to stimuli. (Derived from OED)

Given that uc:Person is a subclass of uc:Animal, this definition entails that a uc:Person is a non-human organism. This problem has now been corrected through UCore's change management process by removing 'non-human' from the definition of 'animal', but further problems remain. (For example Alert Event is treated by UCore 2.0 is a sibling, rather than as a child, of Communication Event; Weather Event, similarly, is treated as sibling rather than as child of Natural Event.

Other examples of UCore 2.0 definitions are:

- *uc:GroupOfPersons* =def A number of people located, gathered, or classed together. (Derived from OED)
- *uc:Organization* =def An organized body of people with a particular purpose, e.g. a business or government department. (Verbatim from OED)
- *uc:PoliticalEntity* =def An organized governing body with political responsibility in a given geographic region. (Derived from OED)

The definition of 'Organization' does not make it clear whether or not organizations are groups of persons. The definition of 'PoliticalEntity' suggests that it should be a subclass of 'Organization', but this is not reflected in the UCore 2.0 taxonomy.

UCore SL, in contrast, rigorously utilizes the structure of the taxonomy in the formulation of its definitions. Every UCore SL term is defined in terms of necessary and sufficient conditions following the Aristotelian schema, which defines each child term 'A' in terms of its immediate parent 'B' together with the differentia 'C' which determines what it is about the B's which makes them A's (as in: a *human* =def. an *animal* that is *rational*). Examples from UCore SL are:

- sl:Government =def. An Organization with political responsibility for governing in a specified GeospatialRegion.
- sl:Organization =def. An Agent that has (1) members which are Agents, (2) one or more Objectives, and (3) MemberRoles (and other AffiliateRoles) which are realized in the pursuit of the Objective or Objectives
- sl:GroupOfPersons =def. A Group that includes only Persons.

The fact that sl:Government is a subclass of sl:Organization is reflected in both the definition and the taxonomy (see table 3). In UCore SL it is possible to state not merely that sl:Organization and sl:GroupOfPersons are distinct, but also that they share no instances in common, since UCore SL includes explicit disjointness axioms.

VI. Current Projects and Future Plans

C2 Core

C2 Core, a DoD-level initiative pursuing C2 data interoperability, is exploring a combined top-down/bottom-up approach, which both extends semantics down from UCore 2.0 while also addressing the bottom-up requirements for information exchange brought by specific user groups. The NCOR team is achieving logical consistency through a top-down extension of UCore 2.0 terms, logically defined using the resources of UCore SL, and applying the result to create a C2 conceptual data model called 'C2 Core'. The latter currently contains over 120 high-frequency terms that define the C2 domain. These terms pertain to situational awareness, structuring a military organization, planning and assigning tasks, decision making, and assessing progress.

Examples of potential targets for extensions of the existing C2 Core include sub-domains such as Strike, Unit Readiness, Planning and Operations, and the Military Decision Making Process (MDMP). Experience in creating UCore SL has yielded a proven process for creating such extensions which results in definitions which are optimized for use both by humans (for teaching and doctrine writing) as well as use by computers (in validation and reasoning).

Using UCore SL to Support Reasoning with UCore Messages

As summarized in [11], we are developing a system which will allow software agents to better understand and reason with UCore-2.0 messaging content in an approach based once again on the logical resources provided by UCore SL. The underlying idea is to treat the XML-labels used in UCore 2.0 messages as annotations for particulars (for instance individual agents) about which these messages contain information. Some particulars are referred to in these messages directly (for instance the military unit that has been given an order to move from place A to place B); others are particulars that must exist for the messages to be correctly interpretable by software agents and whose existence can thereby be indirectly inferred. To make such inferences XML-labels are mapped to ontologies based on UCore SL. Depending on the quality of the mappings, and the quality of the associated ontologies, more and better inferences can be made about the portion of reality described in the messages.

We are working on a method to quantify the quality of these mappings and the ontologies in such a way that we can demonstrate that one ontology is to be preferred over another, or that one mapping to an ontology is to be preferred over another mapping. By using such quantified measures, we can engineer an evolutionary improvement of ontology resources, which can be used across the entire domain of messaging in areas such as C2, where tight integration of messages deriving from disparate sources is required.

UCore SL as Basis for a Cyberwarfare Operations Ontology

While standard military operations doctrine is thoroughly documented in Joint Publications (JPs), Field Manuals (FMs), and other reference materials, this is not the case for military operations in cyberspace. Now, however, with the increasing importance of cyberwarfare, there is a need for standardized terminological resources which can serve as the basis for formulation of sound doctrine and also be applied to other purposes such as the development of international law pertaining to cyberwarfare. Doctrine and law can be written only if experts agree on the semantics of the domain [12]. Drawing, again, on our experience with UCore SL, we propose to identify the semantic content pertaining to cyberwarfare, defining and establishing relations between the high-frequency terms that are common to the relevant subject matter experts. We will then use UCore SL as basis for a Cyberwarfare Ontology, proposing to UCore 2.0 additional terms for inclusion as necessary.

TABLE III: UCORE SL TAXONOMY

sl:Entity		sl:Event
sl:InformationContentEntity	sl:Infrastructure	sl:Act
sl:Analysis	sl:Materiel	sl:ActOfCommunication
sl:Objective	sl:Consumable	sl:ActOfHumanitarianAssistance
sl:ObjectiveSpecification	sl:Organization	sl:ActOfObservation
sl:Opinion	sl:Government	sl:CriminalAct
sl:Plan	sl:PhysicalObject	sl:ImmigrationEvent
sl:TaskSpecification	sl:LivingThing	sl:LawEnforcementEvent
sl:PhysicalEntity	sl:Animal	sl:TerroristAct
sl:Agent	sl:Person	sl:CyberSpaceEvent
sl:Artifact	sl:InfectiousOrganism	sl:Danger
sl:ArtificialAgent	sl:MicroOrganism	sl:Disaster
sl:Equipment	sl:Plant	sl:EconomicEvent
sl:Facility	sl:Vehicle	sl:FinancialEvent
sl:Sensor	sl:SpaceRegion	sl:EnvironmentalEvent
sl:Environment	sl:Property	sl:Epidemic
sl:GeographicFeature	sl:Capability	sl:EvacuationEvent
sl:GeospatialBoundary	sl:PhysicalProperty	sl:HazardousEvent
sl:GeospatialRegion	sl:AtmosphericProperty	sl:Incident
sl:AdministrativeDivision	sl:GeographicProperty	sl:InfrastructureEvent
sl:ControlFeature	sl:OceanographicProperty	sl:MigrationEvent
sl:CoverageFeature	sl:SpaceEnvironmentProperty	sl:MilitaryEvent
sl:GeopoliticalEntity	sl:Role	sl:MissileLaunchEvent
sl:Route	sl:AffiliationRole	sl:NaturalEvent
sl:Track	sl:AgentRole	sl:AtmosphericEvent
sl:Group	sl:CargoRole	sl:GeographicEvent
sl:GroupOfOrganizations	sl:ControlFeatureRole	sl:NaturalEvent (cont.)
sl:GroupOfPersons	sl:ControlledSubstanceRole	sl:OceanographicEvent
sl:InformationBearingEntity	sl:InformationSourceRole	sl:SpaceEnvironmentEvent
sl:Database	sl:MaterielRole	sl:PlannedEvent
sl:Datafile	sl:WaypointRole	sl:PoliticalEvent
sl:Document		sl:PublicHealthEvent
sl:Program		sl:SecurityEvent
sl:Website		sl:NationalSpecialSecurityEvent
		sl:SocialEvent
		sl:StructuralCollapse
		sl:Task
		sl:TransportationEvent

Crosswalks between UCore SL, DOLCE, and SUMO

As part of the design process, UCore SL has a built-in crosswalk between UCore SL and BFO. In an effort to make UCore as widely applicable as possible, additional crosswalks to the other major upper-level ontologies DOLCE [13] and SUMO [14] will be created in order to leverage the knowledge sources that utilize these artifacts.

UCore 2.0 andJC3IEDM

The Joint Consultation, Command and Control Information Exchange Data Model (JC3IEDM) is a model that aims to enable international interoperability of C2 information systems at all levels in order to support multinational (including NATO) combined and joint operations and the advancement of digitization in the international arena [15]. Recognizing that the integration of UCore 2.0 with relevant portions of JC3IEDM would represent a significant interoperability gain, and being aware also that a direct mapping between UCore 2.0 and JC3IEDM is likely to produce inexact results, NCOR is exploring the option of use the logical resources of UCore SL to build semantic bridges between the two resources.

VII. Conclusion

UCore SL, an ontology-based supporting layer for UCore, is designed to work behind the scenes in UCore 2.0 application environments as a logical supplement to the UCore messaging standard. UCore SL builds upon previous work in the biomedical domain on creating consistent extensions on the basis of a common core ontology in order to serve interoperability. UCore SL provides the logical resources for the UCore initiative to do this work.

UCore SL is currently in the beta phase of development, with several current and potential users who are testing it in their application environments and providing valuable feedback in order to help improve future versions of UCore-SL. In order to demonstrate the true value of UCore SL it is necessary to develop a significant user community around UCore SL, one where multiple extension ontologies are subjected to rigorous logical analysis and testing, linked together in computable ways, and used to annotate large quantities of data. In this way it will be possible to show how UCore SL's added logical resources can meaningfully advance UCore's information sharing goals.

Acknowledgements

Thanks are due to Paul Birkel, Brian Haugh, William Mandrick, Randall Dipert, Leo Obrst, Ron Rudnicki, and especially to Deborah Nichols, for valuable contributions to the development of UCore SL.

References

[1] http://ucore.gov/.

[2] http://www.gcn.com/Articles/2009/06/15/Reality-Check-commentary-UCore-info-sharing.aspx.

[3] The Gene Ontology Consortium (2000). "Gene Ontology: tool for the unification of biology," *Nature Genetics*, 25: 25–29.

[4] Amy Adams, "Your database is talking; is anybody listening? Biological linguists develop new ontologies for richer, and cross-database searches," *The Scientist* 2005, 19(17):26.

[5] Barry Smith, Michael Ashburner, Cornelius Rosse, *et al.*, "The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration," *Nature Biotechnology*, 25 (11), November 2007, 1251-1255.

[6] http://www.ifomis.uni-saarland.de/bfo.

[7] http://www.ifomis.org/bfo/users.

[8] Barry Smith, Werner Ceusters, Bert Klagges, *et al.*, "Relations in biomedical ontologies," *Genome Biology*, vol. 6, 2006, R46, 2005.

[9] Army Net-Centric Data Strategy (ANCDS). http://data.army.mil/.

[10] Winters, Leslie and Andreas Tolk, "C2 Domain Ontology within Our Lifetime", *Proceedings of the 14th International Command and Control Research and Technology Symposium (ICCRTS)*, June 15-17, 2009, Washington, DC. http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA503107&Location=U2&doc=GetTRDoc.pdf.

[11] Shahid Manzoor, Werner Ceusters and Barry Smith, "Referent Tracking for Command and Control Messaging Systems", *Proccedings of the 3rd Conference on Ontology for the Intelligence Community*, George Mason University, Fairfax, VA, October 21-23, 2009 (in this volume).

[12] Randall Dipert, "Ethical and Ontological Aspects of Cyberwarfare", to be presented at International Society for Military Ethics Conference, University of San Diego, January 26-29, 2010.

[13] http://www.loa-cnr.it/DOLCE.html.

[14] http://www.ontologyportal.org/.

[15] http://www.mip-site.org/035_Public_Organ_DMWG.htm.