Position Paper for the International Semantic Web Workshop (2001):
Building a Semantic Web for the Intelligence Community

Brian Kettler[1]
ISX Corporation
bkettler@isx.com

Intelink, the U.S. Intelligence's Community's classified intranet, shares many of the challenges of the World Wide Web. Users of Intelink are diverse (e.g., analysts, warfighters, policymakers) and distributed over the world. Information is still largely in English text (and HTML) documents on web servers with keyword-based search tools and browsers being the primary user applications. Security issues are paramount in a community that has traditionally lacked the culture and the infrastructure to share information widely beyond those who have a demonstrated "need to know". Despite the existence of community standards, Intelink producers are in the nascent stages of using HTML metadata tags and content-based markup languages such as XML. The latter presents an additional challenge of gaining agreement among user communities on sets of tags whose meaning is explicit and exploitable by human users *and* software agents.

The Horus Project is a joint effort by the Dept. of Defense Advanced Research Projects Agency (DARPA) and the Intelink Management Office (IMO) to bring semantic web technologies to Intelink and the Intelligence Community. In its second year, Horus is currently refining a toolkit to bring semantic web tools to user sites on Intelink. This will support users in building enhanced, web-based knowledge portals that provide access to both structured data in databases and unstructured data extracted from web sources.

The focus of Horus is to enable and exploit *semantic*-based markup of sources to promote information discovery and integration – ultimately by software agents as well as humans. Users and agents will access, manipulate, and create knowledge that is organized as Horus "knowledge objects". These (conceptual) objects represent real-world entities such as military units, terrorist organizations, and geopolitical events. Information in knowledge objects is linked to its source (i.e., a database or web page). This supports the maintenance of information pedigrees and drilldown to the original sources. User sites will build portals to provide access to these objects, resident in a Horus Knowledge Base (KB).

In addition to working with the site developers of user portals, we are also coordinating with the Joint Intelligence Virtual Architecture (JIVA) Project's Knowledge Map effort, IMO efforts for tagging standards (metadata, security, and content), and related Intelligence Community efforts for document markup (using XML, etc.). We view Horus' approach as unique and complementary to those efforts.

To provide access to information from web sources and databases, the Horus Project is applying the DARPA Agent Markup Language (DAML). Our previous work applied the Simple HTML Ontology Extensions (SHOE) language from the Univ. of Maryland. We building five classes of tools in the Horus Toolkit:

---

[1] Dr. Brian Kettler is the Technical Lead for the Contractor Team for the Horus Project. This team consists of BBN, ISX, and Logical Sterling Federal Systems. *The views in this position statement are those only of the author and do not necessarily reflect the views of the Horus Project's sponsors – DARPA and IMO.*

- ♣ **Ontology Tools** –facilitate the authoring, application, and maintenance of multiple, distributed, lightweight domain ontologies. Our ontologies are written in DAML and leverage DAML Program's emerging tools for ontology authoring and validation.
- ♣ **Markup Tool**s –facilitate the use of domain ontologies to mark up documents in DAML. Our markup tools support manual markup and automated markup. The latter currently uses Lockheed's AeroText product for rule-based parsing of natural language documents into structured entities.
- ♣ **Knowledge Base (KB) Tools** –support the storage of information from marked up documents, database extracts, and user-created assertions about knowledge objects. Currently we are using an Oracle DBMS and the Parka Knowledge Base Management System, a frame system developed originally by the Univ. of Maryland that supports inheritance-based inferencing.
- ♣ **Data Source Access Tools** –support access of online databases to extract information (as DAML) for the Horus KB and to drill down to details.
- ♣ **Portal Building Tools** –support the development of web-based, knowledge portals for users to access and manipulate knowledge objects in Horus KBs.

Horus has been an interesting "use case" for semantic web technologies thus far. We have encountered a number of issues in applying these technologies to a dynamic, real-world domain.   These include:

- ♣ how usable are complex markup languages such as RDF and DAML+OIL, both for ontology definition and content representation.   Right now the tools lag behind the language development.
- ♣ how much to mark up in web sources: i.e., is the goal to duplicate all of the document's content or merely provide a more semantically-rich index to that document
- ♣ how can the subjects of statements in document markup be matched with equivalent entities (e.g., a particular military unit or terrorist event) in information extracted from databases – especially given inconsistency in naming of entities, differences in levels of abstraction, differences in URLs between classified networks, etc.
- ♣ where does markup come from – e.g., how can existing HTML and XML markup be converted to DAML markup automatically
- ♣ where do ontologies come from – e.g., how can existing database schemas, XML DTDs, and other domain models/taxonomies be leveraged
- ♣ how to maintain historical information from data sources
- ♣ how to distribute Horus KBs over Intelink to agents, crawlers, etc.
- ♣ how to support the security/dissemination restrictions on the various versions of Intelink – e.g., (semi-)automated tools for restricted-word search in markup

We expect to encounter additional issues as we put in place the first "Horus-enabled" end user site on Intelink by Fall, 2001. We are coordinating with the integration contractor and principal investigators in the DAML Program to facilitate bi-directional technology transfer between Horus and DAML. We are applying tools from the DAML, XML, and RDF communities to Horus. To the DAML Program and related efforts, we are providing a source of real-world requirements and an "alpha test bed" for Semantic Web concepts and technologies in a dynamic and important domain.