ArdenML: The Arden Syntax Markup Language
(or Arden Syntax: It's Not Just Text Any More!)

R. Matthew Sailors, PhD
University of Texas-Houston Health Science Center
Medical School, Department of Surgery, Houston, TX 77030

ABSTRACT
It is no longer necessary to think of Arden Syntax as simply a text-based knowledge base format. The development of ArdenML (Arden Syntax Markup Language), an XML-based markup language allows structured access to most of the maintenance and library categories without the need to write or buy a compiler may lead to the development of simple commercial and freeware tools for processing Arden Syntax Medical Logic Modules (MLMs)

LEVELS OF ENCODING
Borrowing the concept of levels of detail from other groups within HL7 (Health Level Seven, Inc), a numeric coding scheme for identifying the level of detail was developed and presented to the Arden Syntax Special Interest Group (Arden SIG). The levels are numbered from 0 to 4 in increasing level of detail.

Level 0
Level 0 encoding simply wraps existing Arden Syntax MLMs in their entirety inside a large CDATA (character data) construct. The CDATA type is used to allow for the commonly used greater than and less than signs (">", "<") that are used as tag delimiters in XML-derived markups.

Level 1
Level 1 ArdenML encodes the Arden Syntax hierarchy of categories and slots. Each category becomes a collection of substructures (the slots) and each slot is represented as a separate CDATA field. It is important to note that at ArdenML Level 1 and higher, the normal category and slot identifiers and all terminators(";","/;",";") have been replaced by tags.

Level 2
Level 2 of ArdenML adds another layer of structure to model. The structures added to the citation and links slots in Arden Syntax Version 2 and addition structuring of the author, specialist, and keyword slots that have been proposed for Arden Syntax Version 3 first appear in ArdenML Level 2 encoding. Additional elements representing many of the coded values within Arden Syntax are also present.

Level 3
Level 3 encoding introduces the structuring of whole statements and control structures. ArdenML encodes the blocks of statements inside control structures as nested constructs with additional nesting of the statement blocks that appear in if-then-else conditional statements or within loops. This representation simplifies the processing of the encoded MLM and is in keeping with the model of adding of additional structure to the MLM as the level of encoding increased. Tags to encode comments, identifiers, referenced MLMs, and tokens used to structure data have also been added in ArdenML Level 3.

Level 4
Level 4 encoding structures Arden Syntax down to the level of operators and operands. A DTD for ArdenML Level 4 has not been developed at this time, as this requires the reimplementation of the entire Arden Syntax BNF (Backus-Naur Form) in XML-based constructs. This was deemed to be unnecessary for this preliminary work product.

CONCLUSIONS
At this time it is not clear if encoding Arden Syntax beyond Level 2 has any real value other than as an academic exercise. However, as XML translation tools begin to multiply, ArdenML Levels 3 and 4 may become practical. ArdenML Levels 2 and 3 also show promise as intermediary forms to be used during development and maintenance of MLMs using componentized development tools.

FUTURE DIRECTIONS
An exploration of the use of XST (XML Stylesheet Translation) to translating MLMs into other program constructs mining/knowledge representations without having to build a traditional compiler is planned.

References

Support for this work was provided by DAMD17-98-2-8002.