Much of the knowledge base that serves as justification for the commitment to a given percept during the interpretation process will come from the experts. However, letter frequency, word- and character-lists from documents such as the Vindolanda ink tablets will provide an invaluable source of information which can be used to generate the statistical likelihood of patterns in language and writing which may appear on the texts. We have taken a new approach to the XML encoding of the Vindolanda ink tablets based on contextual encoding (Hippisley 2005). The Vindolanda ink tablets have been encoded with EpiDoc standard XML to a very detailed granularity. The contextual encoding which is then imposed on the documents consists on encoding words, person names, geographical place names, calendar references and abbreviations. For example any instance of the word pulli (='chickens') in a document will be encoded <w lemma="pullus" n="1">pulli</w>. This encoding provides us with the information that the word pulli has the lemma pullus under which we can index this instance of the word and that this is the first instance of this lemma in the document. This information has been used to generate word frequency lists and is extremely useful as a part of a knowledge base to build the ISS on. Further knowledge bases will be generated from the marked up dataset, to provide uncertainty and character frequency lists. Additionally, further work will be undertaken with the experts to generate lists of common percepts and interpretation making processes.  By encoding these in XML, the knowledge sets for the system will be in place.

*Hippisley, D. (2005) "Encoding the Vindolanda tablets: an investigation in contextual encoding using XML and the EpiDoc standards." MA Dissertation submitted for the MA in Electronic Communication and Publishing, School of Library, Archive and Information Studies, UCL.*