Diacritics and Special Characters in EAD

The EAD DTD supports a subset of the ISO8879 standard for characters. The supported sets are documented in the EAD Application Guidelines (version 1.0, section 6.5.2 on General Entities). They include the character entity sets covering: Latin-1 and Latin-2, Greek symbols (also alternate Greek symbols, Greek letters, and Monotoniko Greek), Diacritics, Numeric and special graphics, Publishing symbols, and the General Technical characters. Physically, these character sets live in files which are named according to the form ISOxxxx.ent where xxxx is the particular set (e.g. ISOlat1.ent or ISOpub.ent). Even though the EAD DTD supports these characters, the display and indexing of these characters depends on the systems which store and display the finding aids.

Indexing of Special Characters in OASIS and HOLLIS

OASIS and HOLLIS strip out diacritic marks and normalize special characters before indexing the finding aids so that searching and retrieval will always work no matter what character-generating capabilities users have on their local computer. The normalization rules follow those of HOLLIS where there is overlap with the ALA character set. OASIS supports the indexing and searching of CJK (Chinese, Japanese, and Korean) characters in EAD finding aids, but currently HOLLIS does not.

Inserting Diacritics and Special Characters into an EAD document

To enter a diacritic or special character into your EAD document, you should avoid using text entities (such as "&copy"). Instead, use the designated Unicode Hexadecimal Reference. For example, to make the copyright symbol appear, insert the reference "©" into your EAD document. Please note that the character entity set references in the EAD DTD do not by themselves make these character entities available for use in EAD instances. You must have these character sets available in your XML system in order to make them work in EAD. Some XML editors contain toolbars for frequently used special characters and symbols, which will automatically associate the required Unicode Hexadecimal Reference with a selected symbol. If the program you are using does not include the character or symbol you are looking for, you will need to find the appropriate Unicode Hexadecimal Reference. A useful list of codes for characters and symbols can be found at: http://www.ramsch.org/martin/uni/fmi-hp/iso8859-1.html.