Data-centric XML

Bits and Pieces

XML 1.1

- originally specified 2004, second edition 16.8.2006
- http://www.w3.org/TR/2001/REC-xmlbase-2001062/
- Rationale: Evolution of Unicode
- Changes:
  - version= changed to 1.1
  - Name production changed to support future Unicode versions
  - Addition of more line end characters: U+0085 (NEL) and U+2028 (LINE SEPARATOR)
  - Control characters (U+0001..U+001F) can now be included in a document, but only by reference; extended control characters (U+0080..U+009F) can now be included only by reference
    - U+0000 still not allowed in documents, line break characters can be included literally
  - Documents SHOULD be in NFC if they are in UTF-*
    - processors SHOULD verify NFC unless document is certified NFC

Namespaces 1.1

- Second edition from 16.8.2006
- http://www.w3.org/TR/2006/REC-xml-names11-20060618/
- Inclusion of errata:
  - several editorial and clarification errata
  - relative references are deprecated as namespaces
  - prefix xml is implicitly bound and may be declared; prefix xmlns is implicitly bound and must not be declared; neither may be used as default namespace; elements must not use xmlns: prefix
  - no colons are allowed in ID, IDREF(S), ... attributes
  - namespace is normalized attribute value of xmlns* attribute
- Namespaces are now identified by IRI references (RFC 3987), not URI references
- prefixes can be undeclared with xmlns:prefix=""
  - undeclaration of xml: and xmlns: not conforming

XML Base

- http://www.w3.org/TR/2001/REC-xmlbase-2001062/

- Issue: resolution of relative URIs (and IRIs) always requires base URI (IRI)

  – base URI understood to be the URI of the document containing the relative URI

  – when copying a document, information about original base URI may get lost

- attribute xml:base can be used to declare base URI explicitly

  – follows element nesting: inner xml:base overrides outer one

  –

xml:id

- http://www.w3.org/TR/2005/REC-xml-id-20050909/
- Issue: ID attributes only recognized when doing validation
  - Some specifications (e.g. SOAP) disallow use of doctype declarations altogether
- attribute xml:id is implicitly recognized as an ID attribute
  - attributes "automatically" get type ID
  - an "xml:id processor" must verify that
    - the declared type of the attribute (if any) is ID
    - the attribute value follows the NCName production
  - it should
    - assure that all IDs are unique (violation is non-fatal error)
    - update the infoset [references] property

Associating Style Sheets with XML documents

- http://www.w3.org/1999/06/REC-xml-stylesheet-19990629
- processing instruction <?xml-stylesheet ...?>
  - must occur in document prolog
- pseudo-attributes: semantics like HTML 4.0 <LINK REL="stylesheet" ...>
  - href: URL for style sheet
  - type: mime type
  - title
  - media: screen, tty, tv, projection, handheld, print, braille, aural, all (default screen)
  - charset
  - alternate: yes|no (default "no"); means same as <LINK REL="alternate stylesheet"...>