Knots in the Net

Steps to Standardization of Electronic Buddhist Texts
Trials and Errors:

- Applying the TEI guidelines
- Setting up a production scheme
- Adaptations of the DTD
Applying the TEI Guidelines

• Why not just *use* them?
  – TEI is like a travel guide, not a catechism

• Deciding on coding practices
  – descriptive vs. normative
  – Selection of ‘tags’ (elements)

• How is this done?
How is this done?

- By looking at the texts
- By trying some alternatives
- By discussing them with colleagues (with you!)
- By processing the texts for the intended targets formats
- *gradual* rather than *sudden*
CBETA markup process

- Semiautomatic tagging with two validation cycles:

  Proofread text → AddPara.pl

  Text with basic markup → Validation and correction

  Footnotes → AddFootn.pl

  Text with variants markup → Validation and correction
大方廣佛華嚴經卷第一
于闐國三藏實叉難陀譯
世主妙嚴品第一之一

A simple example:
A simple example (2):

大方廣佛華嚴經卷第一

【明】【三】【宮】【*】

于闐國三藏沙門法師實叉難陀

奉制

譯

于闐國三藏

實叉難陀

(法師)奉制
Some elements added for CBETA

- A specialized milestone element `<JUAN>`
- A group of elements for sound glosses `<FAN>` `<ZI>` `<YIN>`
What else?

- Refinement of the DTD
  - TEIlite + (local adaptations) -
    (unused elements) = CBETA.DTD
  - CBETAXML.DTD using a method introduced by Gary F. Simons
    (see http://www.sil.org/silewp/1998/006/)

- Publication in various formats