



Text Encoding Initiative (TEI)

Stephen Miller

AAC / ÖAW

Texts & Files WS 2006/07

Universität Wien



Oxford Text Archive

- Personal initiative of Lou Burnard (OUCS)
- Initial (and continuing) policy of accession: the “dustbin” approach
- Most texts tagged in OCP COCOA format
 - derived (“diverted”) from print publication projects
- Creation of etexts *incidental*



SGML ISO 8879 (1986)

- Development driven by
 - AAP: American Association of Publishers
 - Pentagon (CALIS project)
- Background in problems and issues of large-scale document provision and maintenance in US corporate enterprises



Keystone Concept

- Descriptive Markup
 - Steven De Rose, “Markup Systems and the Future of Scholarly Text Processing,” *Communications of the ACM*, 30 (1987), 933-47.



TEI Sponsors

- Association for Computers and the Humanities
- Association for Computational Linguistics
- Association for Literary and Linguistic Computing



TEI Funding

- US National Endowment for the Humanities
- Directorate General XIII of the Commission of the European Communities
- Andrew W. Mellon Foundation
- Social Science and Humanities Research Council of Canada



“Poughkeepsie Principles” Vassar College, NY (1987)

- Guidelines principles:
 - suffice to represent the textual features needed for research;
 - be simple, clear, and concrete;
 - be easy for researchers to use without special-purpose software;
 - allow the rigorous definition and efficient processing of texts;
 - provide for user-defined extensions;
 - conform to existing and emergent standards.



TEI Guidelines (P3)

- TEI reported in 1994 with:
 - *Guidelines for Electronic Text Encoding and Interchange (TEI P3)* (Chicago & Oxford: ACH-ALLC-ACL Text Encoding Initiative, 1994)
 - Set of DTDs
 - “Chicago Pizza Model”



TEI Tag Sets

- Prose
- Verse
- Drama
- Speech Transcriptions
- Print Dictionaries
- Terminological Databases



“Future Developments”

- Linguistic and Grammatical description
- Historical Documents
- Manuscript Study & Description
- Base tag sets for “further document types”



TEI Homepage

- <http://www.tei-c.org>
- Teach Yourself TEI Lite
 - http://www.tei-c.org/Lite/teiu5_split_en.html

