Digital Humanities Observatory
Ireland's window on humanities e-scholarship

A project of the **RIA** ROYAL IRISH ACADEMY
ACADAMH RÍOGA NA HÉIREANN

Why not encode?

Kevin S. Hawkins

k.hawkins@dho.ie

08.04.2010 • National University of Ireland, Galway

~ FUNDED UNDER PRTLI . CYCLE 4 2008 - 2011 ~

HEA
Higher Education Authority
An tÚdarás um Ard-Oideachas

NDP
National Development Plan 2007 - 2013

EUROPEAN UNION
STRUCTURAL FUNDS

What are you trying to accomplish?

You want to reproduce the document online to increase access to it. But what do people need to be able to do with that document?

- Search metadata only?

- Look at page or manuscript images?

- Search the full text?

- Read a transcription?

- Perform complex searches?

How would you fill this in?

Desired product	Type of digitisation	What type of content is this appropriate for?
Metadata and page images		
Full-text search (but still showing users the page images)		
Full-text search with words highlighted on page images		
Transcription of the text		
Ability to do complex searching		

A continuum of costs

less expensive

Desired product	Type of digitisation
Metadata and page images	Scanning
Full-text search (but still showing users the page images)	OCR
Full-text search with words highlighted on page images	OCR with coordinates
Transcription of the text	Corrected OCR or keyboarding, plus basic structural markup
Ability to do complex searching	Corrected OCR or keyboarding, plus deep encoding

more expensive

'Levels' of encoding

An approach to levels of encoding has been developed by librarians in the US:
http://purl.oclc.org/NET/teiinlibraries

Non-librarians have remarked that they have found the approach to text encoding illuminating.