

# Digital Humanities Observatory

*Ireland's window on humanities e-scholarship*

A project of the  ROYAL IRISH ACADEMY  
ACADAMH RÍOGA NA HÉIREANN

## Digital workflows

Kevin S. Hawkins

k.hawkins@dho.ie

08.04.2010 • National University of Ireland, Galway

© 2010 Royal Irish Academy



# Your time is not well spent transcribing text

Even if you don't plan to provide page images to users, scanning page images and doing OCR might be a first step. You might even be able to generate basic markup (like paragraphs and page breaks).

Vendors do scanning and sometimes OCR for cheaper than you'll be able to do it in-house.

Could you outsource 'keyboarding' of text to a vendor?  
Or even basic tagging?

Vendors demand unambiguous instructions on how to encode everything they might run into, so you can't just tell them you need 'TEI'.

# What makes a successful encoding project?

- planning

Do document analysis and consult with others before beginning so you don't change your approach half way through a project.

- consistent encoding and metadata creation practice

- quality control (QC) methods

- good workflows

- project-specific documentation

So you remember what you decided and so others can do the work besides you.

# Planning

For a text encoding project, *document analysis* is an important step.

What kind of textual features are present in your source documents?

What do you want to represent?

How should you represent these features?

# Consistency

It's more important to be consistent in your encoding practice than to do it in the canonical TEI way.

If need be, you can go back and change the way you've done things.

Don't encode the same thing two different ways, and don't encode different things the same way.

# Quality control (QC)

Any workflow—even a nearly automated one—needs some form of ‘spot checking’ to make sure that work is proceeding at an acceptable level of quality.

Proportion of content to review—and depth of review—depend on past error rate and on how much is at stake with errors.

# What makes a good workflow?

- As much of the work should be automated as possible.  
You lessen the chance of human error, especially with repetitive tasks.
- If an automated process fails, try to fix the process instead of doing the work manually just that one time since the ‘exception’ will probably occur again.
- Anyone should be able to figure out where someone else left off and pick up the work from that point.  
The sequence of steps—and the stage at which a document sits in the workflow—should be clear. You could use numbered folder names, spreadsheets, or a CMS.



# Documentation

Few people like to write documentation, and most projects find it difficult to keep it current.

For an encoding project, there are so many policies to set that documentation is essential to ensure consistency of encoding over time and with different staff members.