Exercise: Document Description

The object of this small group exercise is to explore the process of deciding on an encoding scheme, and to document the decisions taken.

You will be divided up rather arbitrarily into small groups, and sent off to discuss the sample pages from Punch. After the discussion, we'll ask one person from each group to report back.

Here's our suggestion of how you might proceed:

- First, and very important, choose someone from the group to take notes. Also choose someone to report back. They may or may not be the same person.

- Next, decide what sort of a digitization project you are. You can choose from the following scenarios:

  - You are **librarians**. Your goal is to preserve the best possible digital version of these priceless print volumes. No other copies of Punch are known to have survived the Cataclysm of 2010 so it's up to you to do it right.
  - You are **social historians**. Your goal is to research into early 20th c. British culture and history as revealed in its surviving media. You're using Punch along with many other such resources.
  - You are **Computational Linguists**. Your goal is to produce sophisticated language understanding systems which can detect patterns of lexical usage in a wide range of materials and automatically extract meaning from running text. Or seem to do so.

- Next look at some of the sample pages, and identify which features on the page are of most importance or interest to you. You might ask yourself the following questions:

  - what are the benefits of making this feature explicit in the markup?
  - what are the drawbacks of leaving this feature unmarked?
  - what's needed to identify this feature and mark it up consistently and reliably? (is it well-defined? can it be done automatically? does it require human intelligence?)

- Keep a list of the features you are planning to markup, perhaps with a description of the uses you envisage for them and some idea about the difficulty of capturing them. This will be useful when...

- About half way through your discussion, a bell will go to indicate that you have just had your budget cut by 50%. Now review your decisions!