Laboratory Task:

1. Introduction
   There are software packages that [try to] automatically extract metadata from web sites. The lab lets you look at a couple of these, with a view to evaluating them.

2. Process
   Work individually

3. Task
   Go to the website for Reggie, a metadata extraction tool. The URL is http://metadata.net/dstc/
   Reggie allows for the generation of metadata using a number of different schemas – Dublin Core, GILS, EdNA. And so on. Using one of these, examine the output for the Monash FIT site. Note that Reggie needs Java to be working – you may have some problems.
   Generate metadata for the same site using the UKOLN DC-dot application at http://www.ukoln.ac.uk/metadata/dcdot/ Compare the output with that from Reggie.
   Look at the Nordic metadata project Dublin Core Template web site http://www.lub.lu.se/cgi-bin/nmde.pl  Does this satisfy DC?
   Search the web for other tools that can automatically extract Dublin Core metadata from a website. Can you know if any of these is any good? How?

Tutorial Task:

In addition to engaging in discussion of this week’s tutorial topics….

4. Thinking about Dublin Core….

   - What are some suitable dates that you might need for a website, and how might you express them as qualifiers in a DC structure?
   - If you were using DC as a basis for a CMS, would it be better to use keywords or controlled vocabulary terms for subject in an intranet?
   - If you were designing a CMS, and you were recording a value such as media type or user, would a pull-down menu be a way to prompt users for data entry? How many pull-down options is best? How many is too many?
   - So if you are working with a CMS for a university, how many metadata fields do you use, and how do you structure them?
   - How much would it cost to add this data?