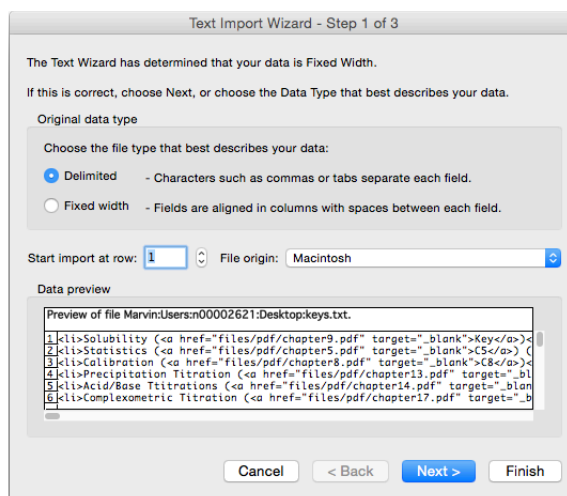# Module 2: Information Science for Chemists (Stuart Chalk)

Assignment Exercise

Now that you know enough about computers and metadata to be information gurus here's an exercise to put your skills to the test. You will create an Excel spreadsheet of chemical property data and metadata.

1. Using the research skills you learned in module 1a find a web page that has a lot of chemical property data on it. It should not be a page that is part of an existing online database, rather a page that is maintained by a science professional(s) as a service to the community. It should also have the following attributes:
   o Useful property data (with units as appropriate) and/or metadata (i.e. descriptors)
   o Chemical names/formulas clear enough to be find CAS numbers and molecular weights
   o Citation(s) indicating the source(s) of the data (i.e. its experimental data)
   o Example pages that meet these criteria are:
     - Common Organic Solvent Properties
     - Amino Acid Properties
     - Thermodynamic Properties of Pure Substances
2. Save the web page as a text file on your computer. Most browsers offer you the option to save as text. This removes any HTML formatting and makes it easier to work with the actual data.
3. Import that text file into Excel. Do this by opening Excel first and then opening the text from the file menu (see right). You should see the dialog box below. In most cases the 'delimited' option works best for importing data. You pick a character on the second dialog page to allow Excel to know how to put the data across different columns (tab, pipe (|), comma, etc.). This will help with organizing the data in the next step.

4. Organize the data that you have imported in columns (different data or metadata) and rows (compounds). Give a title to each column and underneath it identify the datatype for the column (text, enum, integer, float etc.). Some columns may be more than one datatype so try and pick the best one - i.e. the one that is most appropriate for all values that might logically be in that column. Include columns that contain the Chemical Abstracts Registry Number (CAS number) and molecular weights for all of the compounds in the table if they are not already part of the data. An example of what this might look like is below.

	A	B	C	D	E	F	G
	Name	Chemical type	CAS Registry Number (L isomer)	Chemical class	Chemical Formula	Molecular weight (g/mol)	pKa1
	varchar(40)	enum('organic', 'inorganic', 'complex ion', 'organometallic', 'polymer')	varchar(16)	enum('strong acid','weak acid','strong base', 'weak base','amino acid', 'solvent', 'aromatic','ionic liquid', 'drug', 'pesticide' ,'herbicide')	varchar(100)	float	float
	alanine	organic	56-41-7	amino acid	C3H7NO2	89.09	2.3
	arginine	organic	74-79-3	amino acid	C6H14N4O2	174.2	1.8
	asparagine	organic	70-47-3	amino acid	C4H8N2O3	132.12	2.1