**Instructions for use of MicroPIE in ETC**

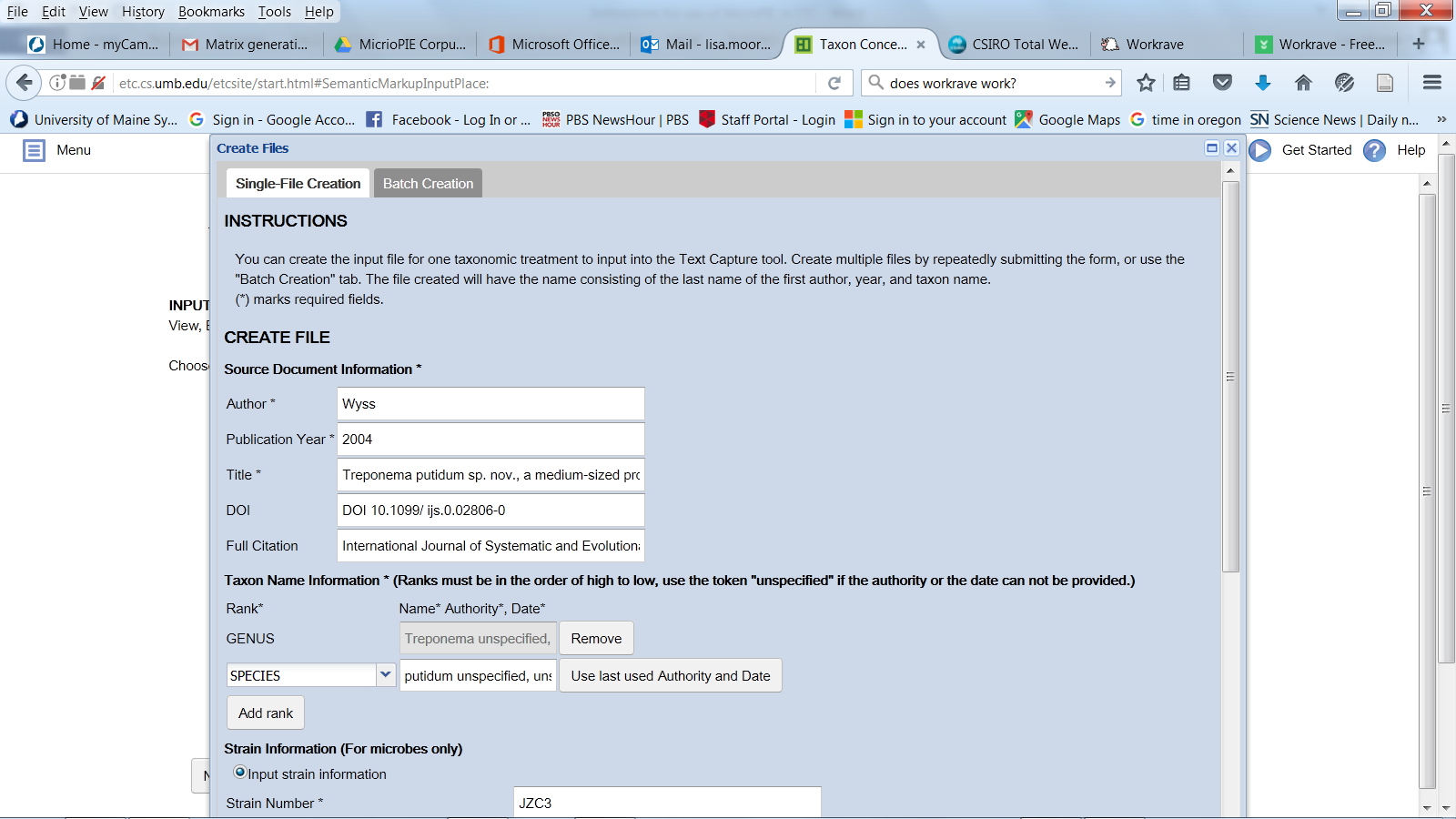
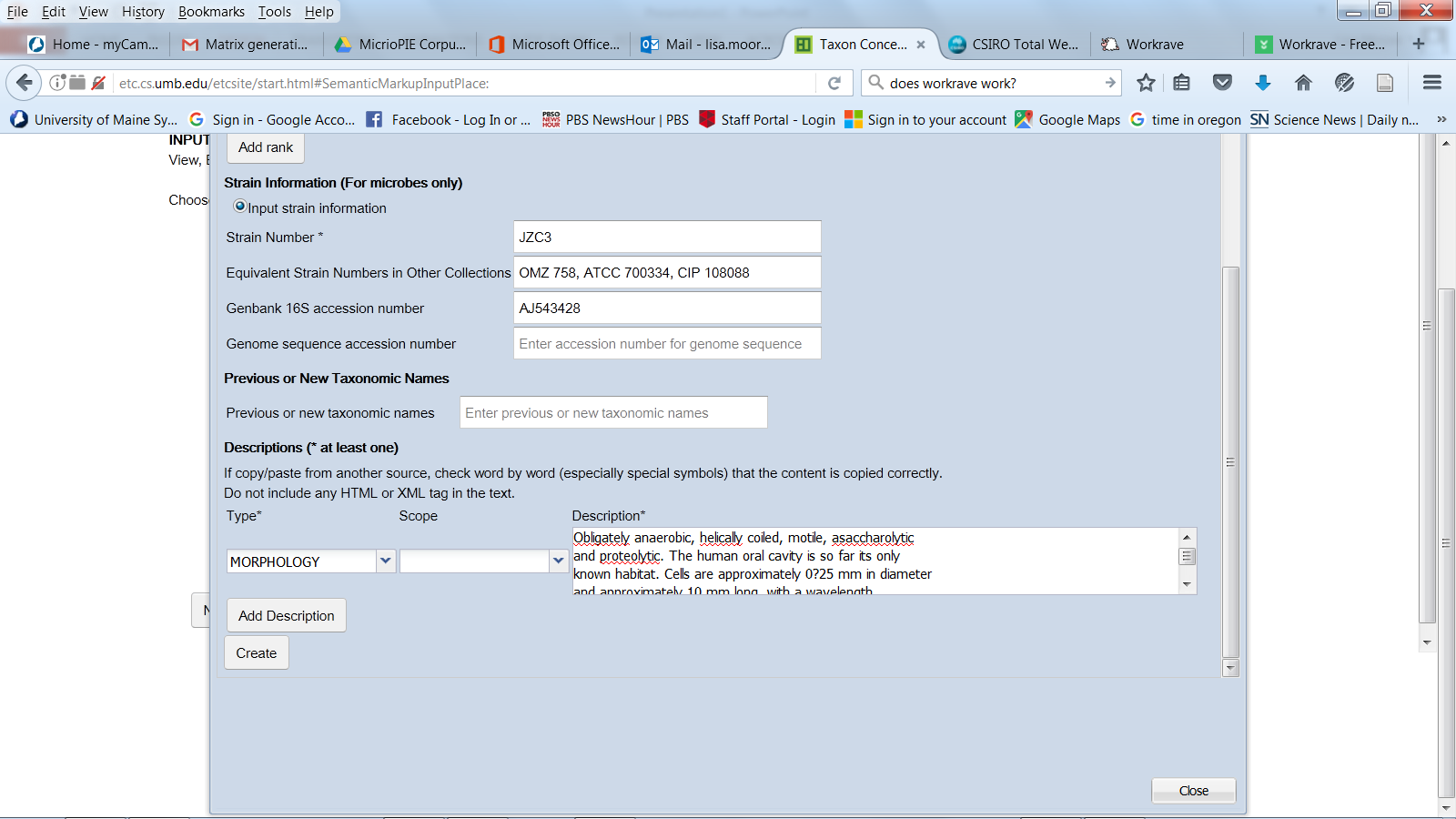
<http://etc.cs.umb.edu/etcsite/start.html>

**1. Select articles of interest**

* Taxonomic description at the ends of articles, such as those published in International Journal of Systematic and Evolutionary Microbiology, were used to develop MicroPIE and work well.
* Unfortunately, the text from the entire article cannot be included as this potentially contains information about other microbes or species/strains. MicroPIE cannot distinguish between the information provided for different microbes.
* Additionally, MicroPIE cannot extract information out of tables or figures at this point.

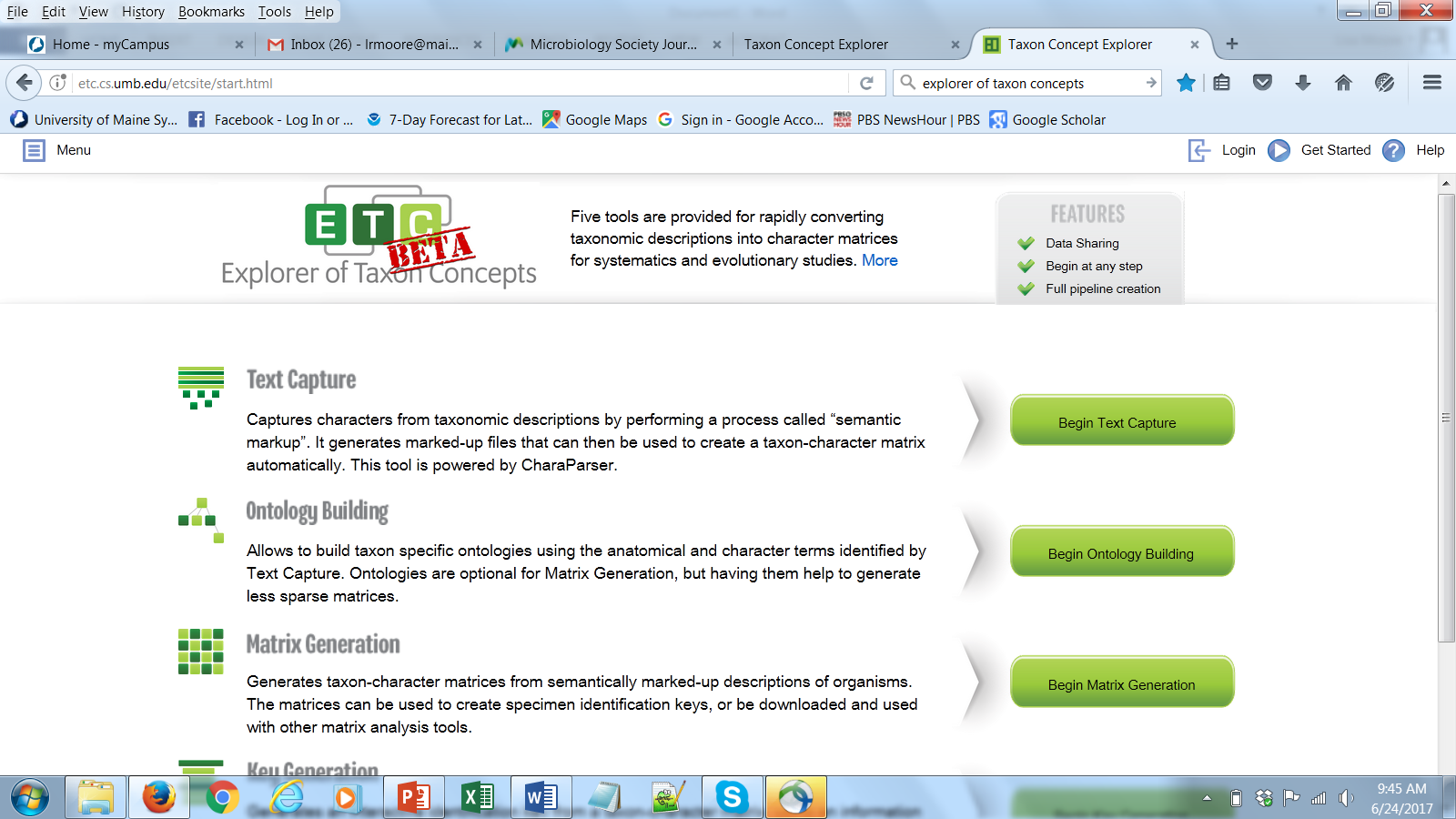
**2. Create Input for Text Capture** tool (this is the trickiest part of the entire process!)

* You can get to this tool in multiple ways, such as through the File Manager or at the beginning of starting Text Capture.
* Create a folder within your Owned folder by clicking on Add Folder.
* The following instructions are for Single File Creation; and screen shot can be seen on the next page.
* Enter the Source Document Information
* Enter the Taxon Name Information
  + First enter genus name followed by unspecified, unspecified. Note that ETC requires Authority, Date information as it is typically used for many other organisms.
    - If your article has a description for the genus, enter this first and do not click on Add Rank or Input strain information
  + Second, enter species name by clicking on Add Rank after entering the genus name. Again, enter unspecified, unspecified for Authority, Date
  + Third, click Input strain information.
    - Under Strain number, enter the Type strain that the taxonomic description refers to. Equivalent Strain Numbers in Other Collections can usually be found at the end of the description, and sometimes at the end of the abstract.
    - 16S rRNA accession number should be for the type strain and is sometimes in figures or tables
* Enter the Description
  + Use Morphology as the type of description; do not add additional types of descriptions for prokaryotes
  + Paste your description from the paper. It is best to copy from the html version of the paper to avoid pdf to text conversion errors.
  + Important! Make sure all errors are corrected *and* it is a single paragraph of text.
* Click Create button at bottom when done.
* Go to File Manager to view the folder and the files. You can edit individual files by clicking on the filename, then clicking on edit.



[Grab your reader’s attention with a great quote from the document or use this space to emphasize a key point. To place this text box anywhere on the page, just drag it.]

**3. Use Text Capture to create xml marked up input files for running through ETC-MicroPIE**



Click on Begin Text Capture

*Step 1: Create Input*

* Use File Manager to create folder for saving files
* Select folder to use
* Click on Next Step in Text Capture

*Step 2: Define Task*

* Give your new project a unique Task Name
* Select Taxon Group – choose Bacteria
* Click on Next Step in Text Capture

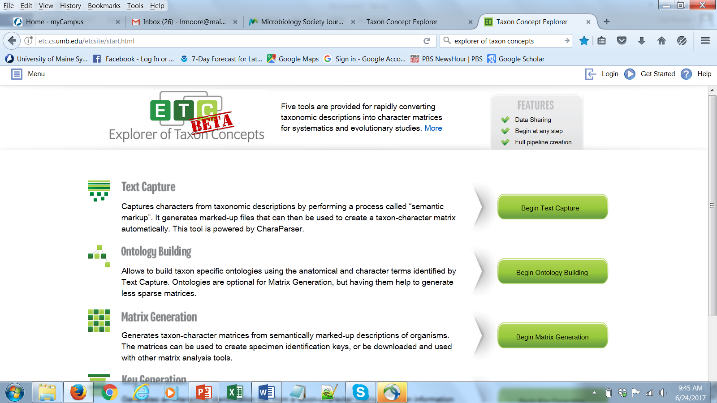
*Steps 3 – 5*: MicroPIE works in background to parse the information

*Step 6*: Click Next Step in Text Capture to finish Parsing

Step 7: Output

* Two output folders are created and put into your designated folder for the task, which will have “\_output\_by\_TC\_task\_” appended to the folder name. An email will be sent to you when the task is finished.
* Go to File Manager to view xml marked up descriptions that can now be used for input

**4. Use Matrix Generation to make taxon-character matrix**

Click on Begin Matrix Generation

*Step 1: Select Input* – select existing folder containing xml marked up input files (this is the output from the Text Capture above, i.e. “\_output\_by\_TC\_task\_”; Click on Next Step in Matrix Generation

*Step 2: Define Task*

* Give your new project a unique Task Name
* Select Taxon Group – choose Bacteria
* Click on Next Step in Matrix Generation

*Step 3: Generate*

Click on "Next Step in Matrix Generation" to see the matrix review page, or Click on Download Matrix Now.

*Step 4: Review*

* There are 3 tabs – Matrix, Annotation, Analyze – that can be used to make corrections and optimize matrix output
* Highlight all taxa that you would like to review at any one time on the screen
* Click whole-organism to review all the characters, or click on individual characters of interest
* You can Export Matrix to get your .csv formatted matrix, or Click on "Next Step in Matrix Generation" and then download your matrix.

Notes:

You can always resume the task from Task Manager. All your work at each step is saved, so you won’t lose what you have done.

Although we have more than 50 characters that can be extracted, the matrix output will only show characters that are extracted.

The characters are listed in alphabetical order from left to right.

There is a way to allow the output matrix to include the sentences from which the values were obtained for a particular character. Though this is a new feature and I need to talk to Hong about how to do it.

Example of output

