

WGBH - DCAMM SCRUM Report 2	
Contact	<p><u>Github accounts</u>: elisa3lopes, rlee99, murtio, carmen-araujo, jenajjedu</p> <p><u>Email addresses</u>:</p> <ul style="list-style-type: none"> <li>• Jena Jordahl <a href="mailto:jenajj@bu.edu">jenajj@bu.edu</a></li> <li>• Elisa Cordeiro Lopes <a href="mailto:elisacl@bu.edu">elisacl@bu.edu</a></li> <li>• Richard Lee <a href="mailto:rlee99@bu.edu">rlee99@bu.edu</a></li> <li>• Murtadha Bahrani <a href="mailto:murtadha@bu.edu">murtadha@bu.edu</a></li> <li>• Carmen Sabrina Araujo <a href="mailto:sabrinaa@bu.edu">sabrinaa@bu.edu</a></li> </ul>
What have I worked on?	<ul style="list-style-type: none"> <li>• Finding key works to extract valuable data from the nested list data. The idea is to extract simple data to be able to do analytics or put in on a pdf</li> <li>• Built a parser that extracts individual project name, project code, and the contractors/companies involved in each. However, this parser misses a few contractors/companies in the listing.</li> <li>• Configuring Grobid on the SCC was not a success.</li> <li>• Tested some code that works with merged cells (Link: <a href="https://github.com/jsvine/pdfplumber/issues/84">https://github.com/jsvine/pdfplumber/issues/84</a>)</li> </ul>
Have I talked to the client recently? When are we meeting with them next?	<ul style="list-style-type: none"> <li>• We talked with the client on Wed 3/11/21 11:30-12:30pm.</li> <li>• Third meeting with the client on Thurs March 18th (or another date since thursday is wellness day)</li> </ul>
What will I be working on next?	<ul style="list-style-type: none"> <li>• Will try debugging pdf parser to make sure it accounts all of the contractors/companies involved in each project.</li> <li>• Merge codes for PDF parsing and run that on all</li> <li>• Meeting with an IT staff to get insight of how we can parse PDF into CSV.</li> <li>• Still finding key words to extract valuable data from nested list</li> </ul>

	<ul style="list-style-type: none"><li>• Updated the csv to list converter to account for numbers with commas</li></ul>
Have I run into any issues? Do I need help?	<ul style="list-style-type: none"><li>• PDF parser misses a few contractors/companies. Need to make sure that all contractors/companies are read.</li><li>• When transforming data from csv to list, using commas as a splitter will divide the numbers containing commas into two</li><li>• Problems when submitting pull request in the required format and committing to new branches</li></ul>