Final Deliverables for WSDOT Project:

My Final Deliverables will be threefold: the beginnings of a Circumstance Catalog, The results of my Public Disclosure Request analysis, and a Report on the Recommended Records for Open Data. I will be working on the first two deliverables in tandem: creating the fields for the Circumstance Catalog and finishing my analysis of my Public Disclosure Records. Once these two are done I can work on my final recommendations.

## What Still Needs to Be Done:

## Circumstance Catalog

1. Create Fields
   1. Will Do Week 6, and can start right away since the criteria will be based on the results of my Coordinator interviews from Week 2-4 and the article *Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework*, the existing WSDOT vocabulary from the DOTs Data Library, and the risk assessment chart used by the WSDOT Open Data Committee.
2. Apply Fields to Top Five Requested Data Types
   1. Will do Week 7 and 8
      1. This cannot be done until text-analysis of the request field is completed, which I am still working on – See PDR Analysis

### PDR Analysis

Good News: Data cleaning is mostly done

Bad News: I’m still struggling to find a method for finding the most frequently requested datasets, which I can learn in the next two weeks. Though I’ve been studying the NLTK Toolbook, I don’t think have the time to apply these skills yet to the dataset. I’ve just figured out how to tokenize the dataset, but that is all I am confident in. I’m thinking of moving away from working with NLTK, and moving into using cruder techniques in Pandas. I can loop through columns looking for strings which match keyword terms or variations on PRD related keyword terms. Luckily, I have a set of keywords from a study done on WSDOT data from I want to test out on the Requested Items field. Through this process I will be able to see how effective these terms are and suggest others. For the Company field, I’d like to use the same process to classify companies by type and subtype.

1. Create keywords for Company and flesh out variations on PRD Keywords
   1. Can start immediately, finish by middle of week 6
2. Create loop functions to find keywords in Company and Requested Items Field
   1. Finish By End of Week 6
3. Create New Columns for Record Type and Company Type
   1. Finish By End of Week 6
4. Deliver Frequencies of Record Types and Company
   1. Finish By End of Week 6
5. Complete Answer Research Questions Using Record and Company Type Column Data
   1. Finish By Middle Week 7

### Report on Recommend Records for Open Data

This will start with the results from my analysis, explain the use of the Circumstance Catalog, present the first five entries of the circumstance catalog as an example to WSDOT, and finally list my recommendations.