Rank likely sensitivity of each document

Create Document-level descriptive statistics, including language,

Types of IIDs found, number of unique IIDs, topic

Blue: physical object

Red: digital process

Green: digital objects

Yellow: Machine learning

Must be used

Purple: machine learning

Could be used

Create table of all the files

Scan for IIDs with Regexes

Scan for Names, Dates?,

Money with NER

Determine content of each

Paragraph/document

Read in files individually &

Classify language. Pass to