Many government documents contain sensitive information that must be identified and protected before the documents can be released to the public. While manually reviewing such documents for sensitive information it can be important to determine contextual information about specific entities that are mentioned in the documents and whether the information that is discussed about these entities is already in the public domain. In this project, you will develop a system that can automatically identify external information about specific entities from publicly available knowledge graphs (e.g. Wikidata or DBpedia). The system should be able to assist human sensitivity reviewers by identifying entities that are referenced by different names in the collection (based on the entity's attributes) and whether personal information about named entities is in the public domain.

You will work with named entity recognition tools (e.g. spacy https://spacy.io/) along with entity linking tool such as ReFinED (https://github.com/amazon-research/ReFinED) or DBpedia Spotlight (https://www.dbpedia.org/resources/spotlight/). A graph databases such as Neo4j (https://neo4j.com/) will likely also be used to dynamically build a definitive view of the entities within the document collection.

- **Summary of what was agreed last week**
  - Attempt second method for entity classification
  - Add analytics to right hand side of document page
  - Add more colour

- **Progress made in the past week**
  - Attempted second classification method, ~0.5 accuracy, not as effective
  - Added initial analytics to right hand side of document view
    - Most frequent entities
    - Documents containing selected entity
    - Most sensitive entities bar plot? (Not implemented)
  - Changed navbar colour schemes and added entity highlighting
  - Continued to improve site flow

- **Main questions for discussion**
  - Side navigation bar appears empty now analytics and document view is combined, unsure what to add to it
  - Additional document analytic features
  - Beginning corpus analytics page

- **Feedback from meeting**
  -