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**
  - Write up classifier experiment
  - Add topic names and related documents to corpus analytics
  - Reduce icons and add document entities on view documents page
  - Potentially introduce an individual page displaying information for each entity

- **Progress made in the past week**
  - Wrote up classifier experiment – currently no conclusion, would rather wait until after feedback on the initial draft
  - Added related documents to corpus analytics page, didn't manage to add topic names yet
  - Reduced icons and added three most sensitive entities on the view documents page
  - Have not implemented a view entity page

- **Main questions for discussion**
  - Feedback on classifier experiment write up
  - Unsure on the working of creating a topic name from a topic
  - Whether a view entity page is necessary or not

- **Feedback from meeting**
  -