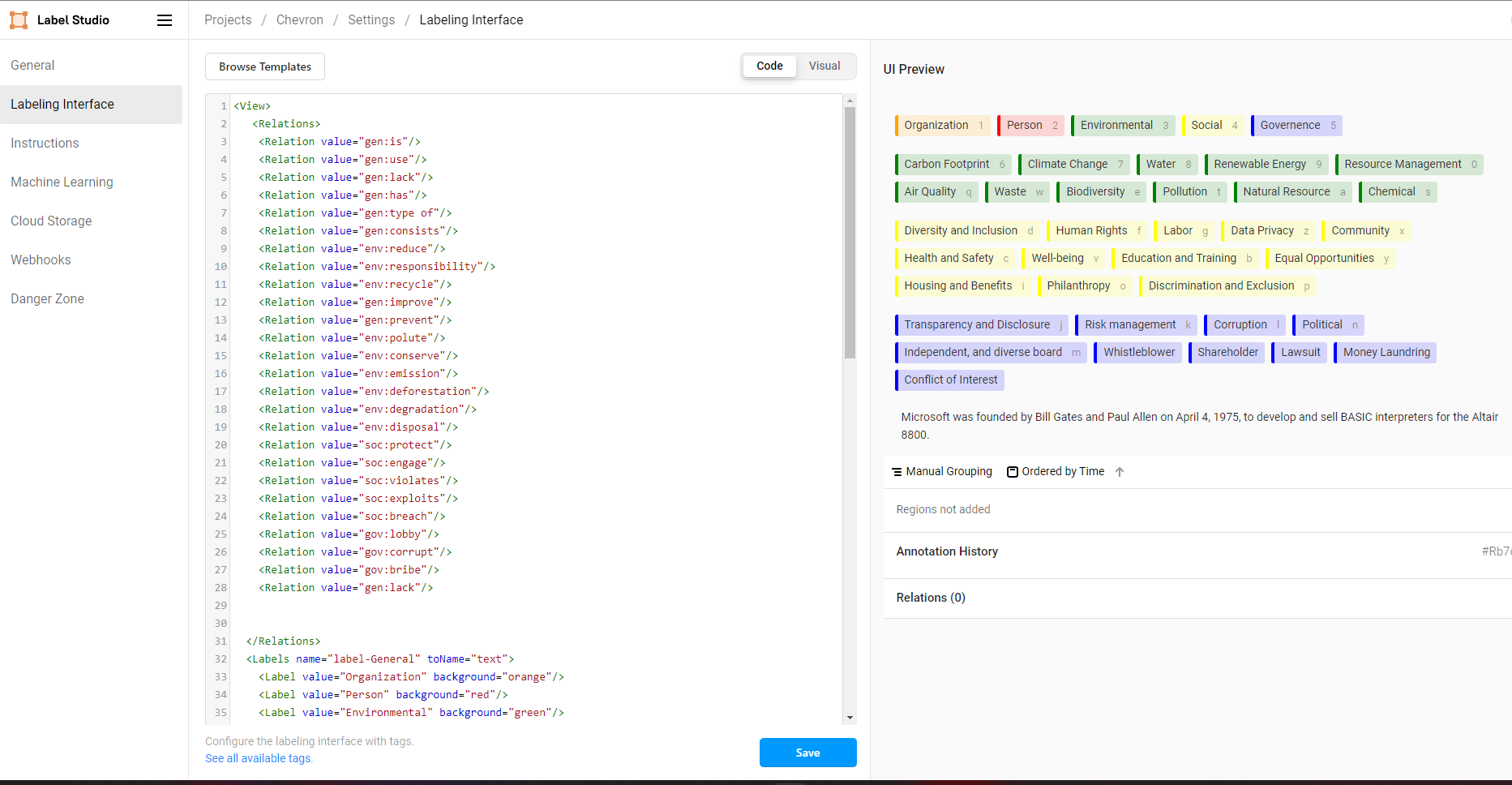
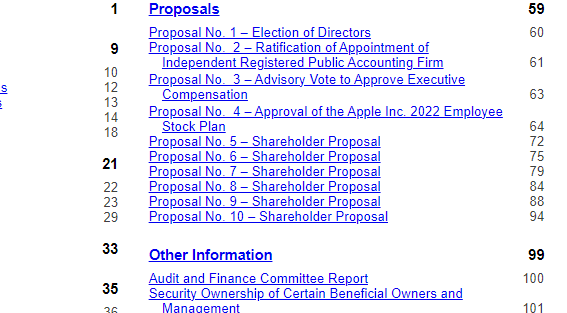
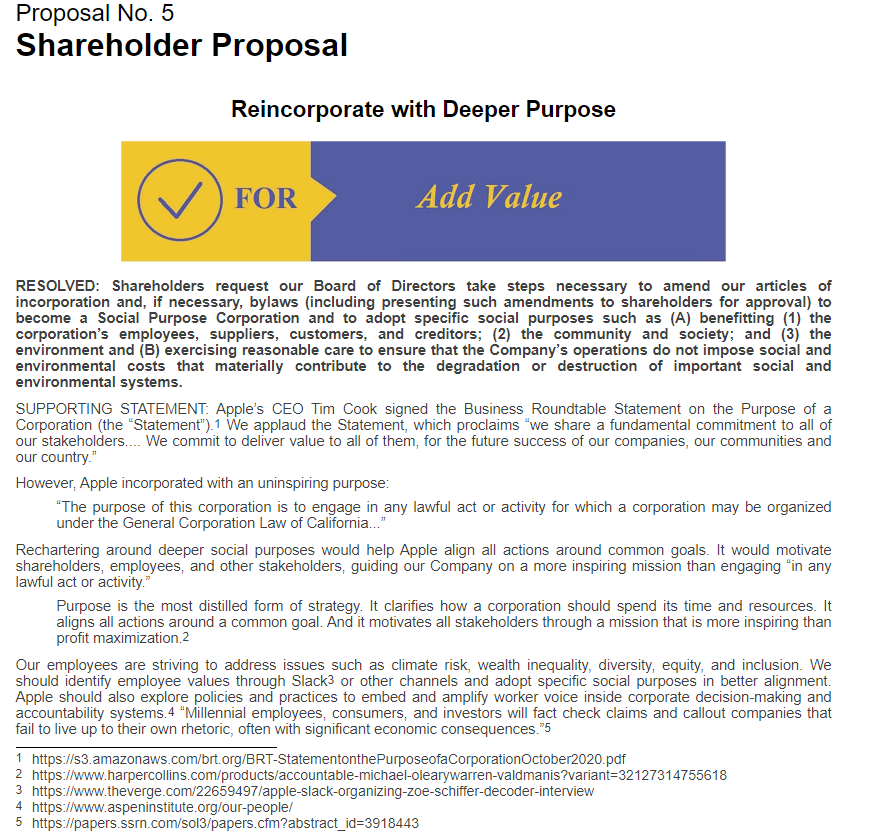
Label Work Flow

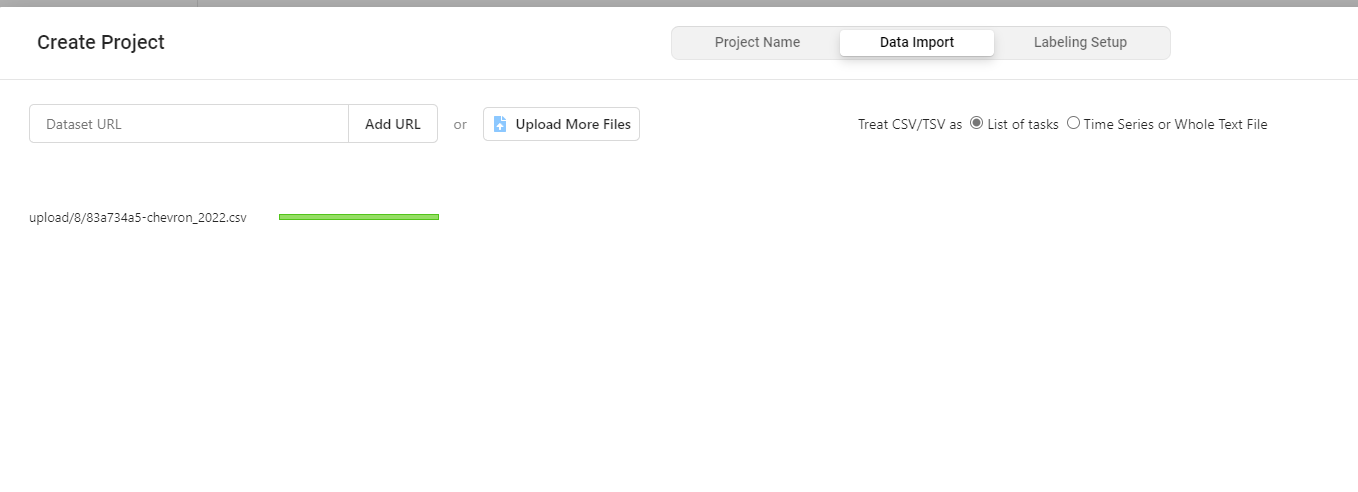
# Getting the tool

1. Install locally: [Label Studio](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi9tYX56qr9AhVxlGoFHT5BC1EQFnoECAkQAQ&url=https%3A%2F%2Flabelstud.io%2F&usg=AOvVaw3Q7TtXagGcNqnEdpC5tTxp)
   1. I would recommend Docker Installation
2. Choose your task: NER, Sentiment Analysis or Relation Extraction
3. Get the schema for your labeling
4. Create your project, you will insert the data later
5. Copy and past it in the Labeling Interface
   1. If you get into trouble, go through the documentation
6. 

# Getting the Data

1. Choose a company that you would like to label and work with
2. Go to one of their DEF 14A documents
3. Look into the table contents, they should be something like this (example: Apple) 4. 4.  
   
4. We are looking for the Shareholder Proposals
5. Here is an example of a Shareholder Proposal  
   
6. You should copy-paste the text into a csv document under a column name: text
7. Take the links (if there are any) of the end of the page and put them into another csv file, or you can put them into a python list.
8. Now we extract the text from the links:
   1. You will notice that some of them are pdfs, those are not required because it will be harder to extract the text, but you are encouraged to do it
   2. For the normal news links, you will use the notebook crawl\_news.ipynb

# Insert your data to label studio

1. 
2. Note: Name your column **text** for your data, I had trouble importing it otherwise.