

 Sora Kanosue
sorakanosue@berkeley.edu

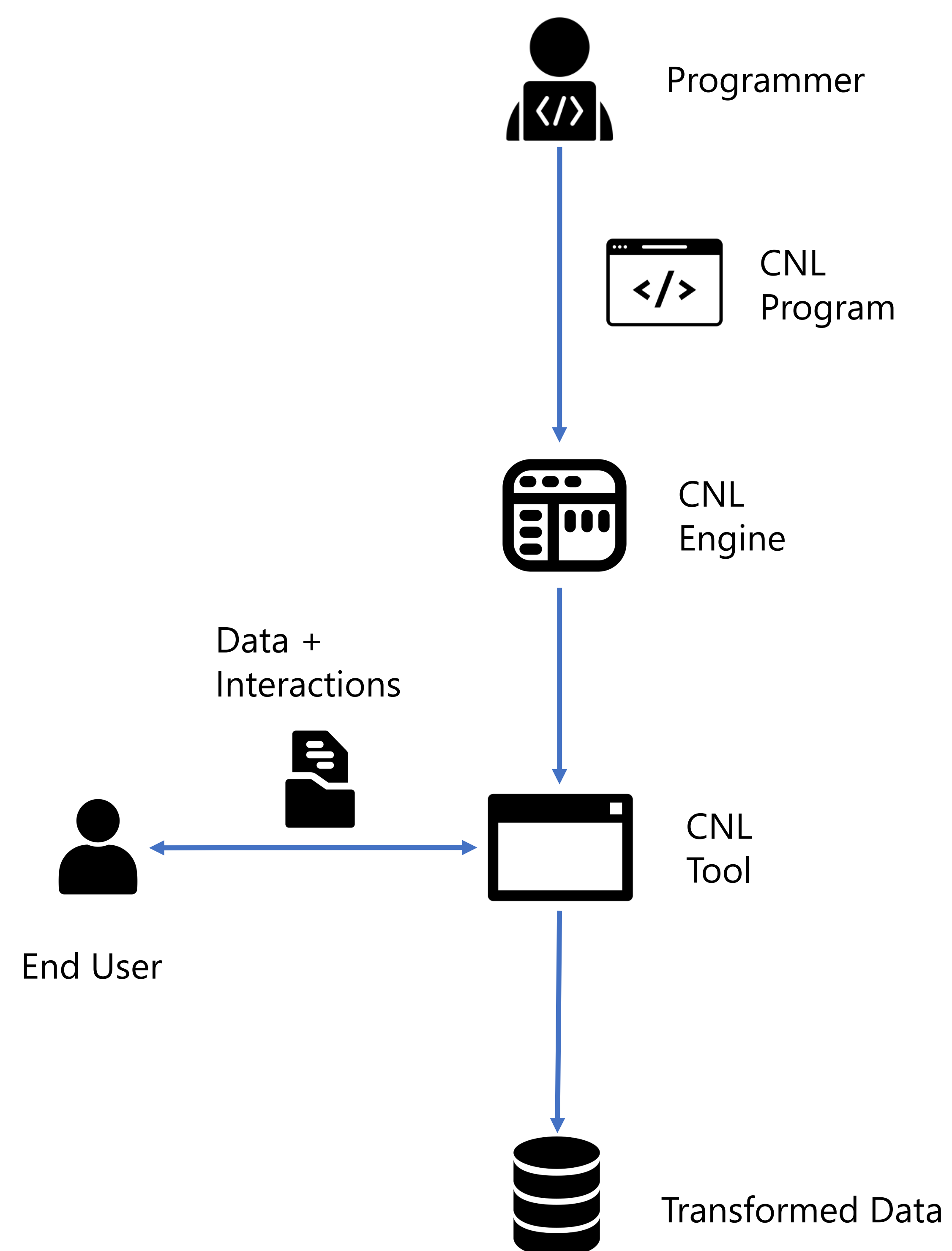
 Sarah E. Chasins
schasins@cs.berkeley.edu

 Parker Ziegler
peziegler@cs.berkeley.edu

Background

- Domain experts in the social sciences often find themselves performing **repetitive tasks restructuring large amounts of data**.
- They often don't have the technical expertise to automate these tasks.
- Writing bespoke tools for these tasks from scratch is **hard**.

System



CNL: A DSL for human-in-the-loop data transformation GUIs

Illustrative Example using CNL:



```
def dataset_upload_stage():
    file_upload_input = FileUploadComponent('csv', "Upload new data sets here:")
    dataset_name_input = UserInputComponent(str, "Enter a name for the dataset you're uploading:")
    def save_dataset():
        saved_table = db_utils.create_table_from_csv(dataset_name_input, file_upload_input)
        return saved_table
    processor = LambdaProcessor(save_dataset)
    my_result = results.Result(processor.result, "Here is the saved dataset: ")
    results.show_results([my_result])

# After defining the stage, we add it to our Tool
tool.add_stage('dataset_upload', dataset_upload_stage)
```

Stages use **Components** to describe GUIs, use **Processors** to describe data transformations, and can show **Results**.

Upload new data sets here:
 No file chosen
Enter a name for the dataset you're uploading:

Here is the saved dataset:

<u>hls_internal_id</u>	First Name	Last Name	Age
1	Oski	Bear	500
2	Carol	Christ	1000

FY2017_Short

FY2018_SHORT

Programs in CNL define **Tools**, each containing multiple **Stages**.