DATA71011: Understanding Data and their Environment

Provenance: Assignment 1

User ID: 11340063

Name: Reethu Biju

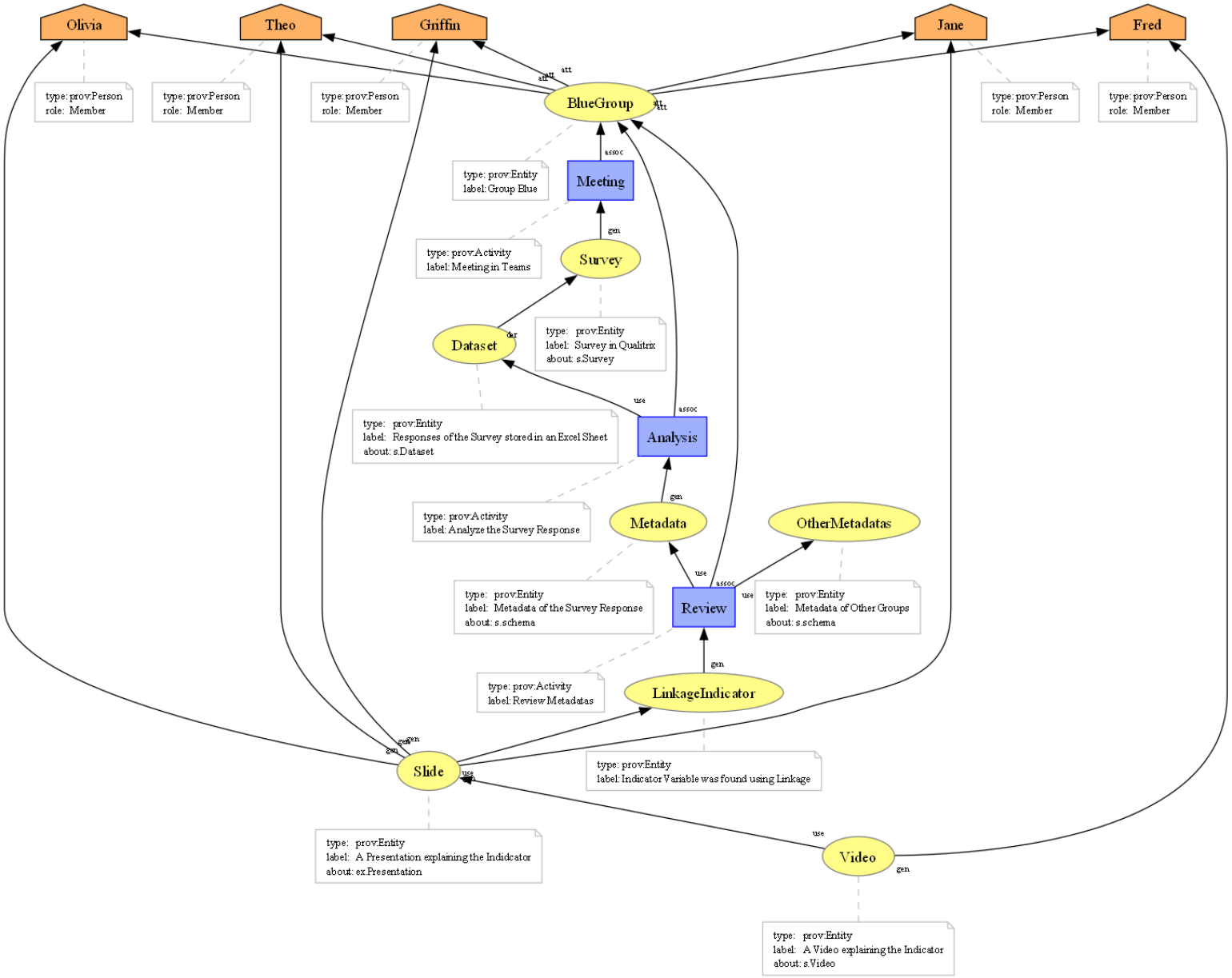

Figure: Provenance Diagram

In the Provenance Graph, the students are defined as agents. These students are attributed to the entity BlueGroup which represents the group. The group was associated in conducting a meeting to generate the survey. The dataset which comprised of the survey responses was then derived from the survey. This dataset was then analyzed by the group to generate the metadata of the dataset. The metadata obtained and metadata from other groups were reviewed to find the Indicator using linkage variables. The founding was then explained in a presentation slide which was done by 4 of the group members and the video was then recorded explaining this presentation by the last group member. The agents were not given their actual names as

identifiers. The activities were linked with the entity BlueGroup instead of each member to reduce the complexity of the graph. This is the reason why activities like meeting, analysis and review was associated with BlueGroup. I have assigned a role for each student as member, because they are part of the same group. From what had happened during week 1-2, I haven't included the part where we were supposed to send the metadata or the video for reviewing to the professor because it would increase the complexity of the graph. Due to this reason agents like professors and activities like receiving feedback was not included for the provenance graph. The graph includes main events that took place and focuses on what the students have done for each week. My main objective on creating the graph was so that anyone would be able to understand on seeing it and hence the diagram has low granularity. This was taken into consideration even while naming the identifiers for entities and activities. To make it more understandable, a label has also been given to each entity and activity describing briefly its role in provenance. If I have to convert the diagram to have higher granularity, I can include events like formation of the group, the allocation of survey topic, distribution of survey links and receiving feedbacks from the professor on the metadata and video submitted. To lower the granularity of the provenance graph I can remove events like attending the meeting, associating each student to the group, reviewing multiple metadata, and creating presentation. The challenging part of writing provenance was deciding on how detailed should I mention the events. This included ensuring that the provenance captured the essential relationships without unnecessary complexity. Another difficulty I found while writing the provenance was not being able to display the label of the relationship. This was because for someone who has no technical expertise would find it difficult to understand the kind of relationship that's being displayed in the provenance graph.