



Universidad Politécnica de Madrid

Escuela Técnica Superior de Ingenieros Informáticos

Data Science Seminars

The Scientific Paper of the Future

Authors:

José Antonio Ruiz Heredia

Teacher:

Daniel Garijo

Date:

March 30, 2025

1 Workflow Description

This document captures the provenance of a survey study, detailing the data, software, and workflow used. The study follows a structured research process:

- **Data:** Survey data collected from two iterations of the survey.
- **Software:** A statistical package for data analysis.
- **Workflow:**
 1. Initial survey design.
 2. Conducting the first survey.
 3. Revision of the survey.
 4. Conducting the second survey.
 5. Compiling and analyzing data.
 6. Publishing the final research paper.

Therefore, a high-level representation of the diagram data flow can be structured as:

1. **Survey Design:** Laura designs the initial survey.
2. **First Data Collection:** Jack and Jill conduct a survey to generate data.
3. **Survey Revision:** Laura revises the initial survey, producing a new updated survey.
4. **Second Data Collection:** Peter and Paula conduct the revised survey to generate new data.
5. **Data Analysis:** Zack compiles the the initial and revised data, analyzes them, and produces the final research paper to be published.

2 Entities

- **Survey_1:** The initial survey designed by Laura.
- **Survey_2:** The survey revised by Laura.
- **Data_1:** Survey data collected by Jack and Jill.
- **Data_2:** Survey data collected by Peter and Paula.
- **Paper:** The final research paper published by Zack, Laura, and other co-authors.

3 Activities

- **Design_Survey:** Laura designs **Survey_1**.
- **Conduct_Survey_1:** Jack and Jill conduct **Survey_1** to obtain **Data_1**.
- **Revise_Survey:** Laura revises **Survey_1** to create **Survey_2**.
- **Conduct_Survey_2:** Peter and Paula conduct **Survey_2** to obtain **Data_2**.
- **Analyze_Results:** Zack compiles and analyzes **Data_1** and **Data_2** to produce the final Paper.

4 Use & Generation

- *Survey_1 WasGeneratedBy* *Design_Survey*.
- *Data_1 WasGeneratedBy* *Conduct_Survey_1* which *used* *Survey_1*.
- *Survey_2 WasGeneratedBy* *Revise_Survey* which *used* *Survey_1*.
- *Data_2 WasGeneratedBy* *Conduct_Survey_2* which *used* *Survey_2*.
- *Paper WasGeneratedBy* *Analyze_Results* which *used* *Data_1* and *Data_2* .

5 Agents

- **Laura:** Designer of *Survey_1* and *Survey_2* and co-author of the final research *Paper*.
- **Jack & Jill:** Conductors of *Survey_1* and attributed to the generation of *Data_1*.
- **Peter & Paula:** Conductors of *Survey_2* and attributed to the generation of *Data_2*.
- **Zack:** Analyst and lead author of the final research *Paper*.

6 Revision & Derivation

- *Survey_2* is *derived from* *Survey_1* through the *Revise_Survey* activity.
- *Data_2* is indirectly *derived from* *Survey_1*, as it is the data collected from the revised *Survey_2*.
- *Paper* is *derived from* *Data_1* and *Data_2* through the *Analyze_Results* activity.

7 Plan

- **Initial Plan:** The initial plan consists of the initial survey design and data collection.
- **Revised Plan:** The revised plan consists of updating the survey and conducting new research.

8 PROV diagram

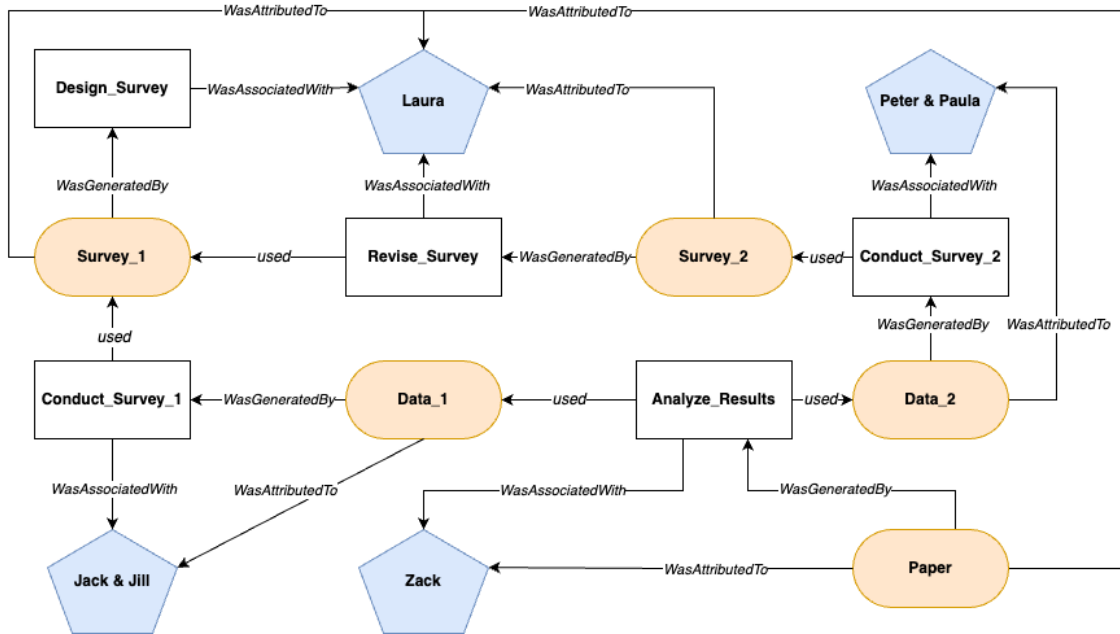


Figure 1: Graph representation of a set of provenance statements