Provocation: Provenance of exploratory data analysis processes are too complex and ad hoc to be useful.

Kai Xu

University of Nottingham, UK

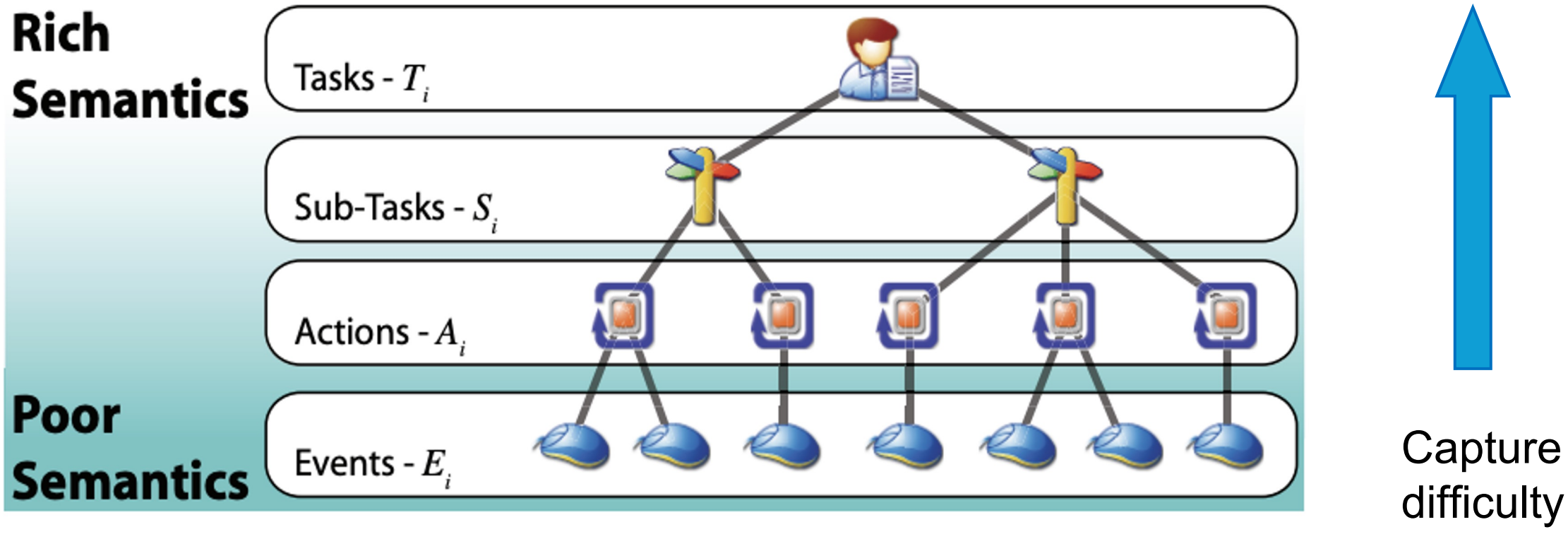
Types of (Analytics) Provenance

- A visualisation perspective

Types of Provenance Information	
Data	The history of changes and movement of data, which can include subsetting, data merging, formatting, transformations, or execution of a simulation to ingest or generate new data
Visualization	The history of graphical views and visualization states
Interaction	The history of user actions and commands with a system
Insight	The history of cognitive outcomes and information derived from the analysis process, including analytic findings and hypotheses
Rationale	The history of reasoning and intentions behind decisions, hypotheses, and interactions

Context for the EDA?

Semantic layers of analytic provenance



Purpose for Provenance

Purposes for Provenance	
Recall	Maintaining or recovering memory and awareness of the current and previous states of the analysis
Replication	Reproducing the steps or workflow of a previous analysis
Action recovery	Maintaining the action history that allows undo/redo operations and branching actions during analysis
Collaborative communication	Communicating and sharing data, information, and ideas with others who are conducting the same analysis
Presentation	Communicating the insights or progression of the analysis with those who are not directly involved with the analysis themselves, such as general public, upper levels of management, or analysts focusing on other areas
Meta-analysis	Reviewing the analytic processes themselves in order to understand and improve aspects of the analysis (such as process efficiency, training efficiency, or analytic strategies)

All useful for EDA?

- For the analysts
- Reproducibility
- Exploration space tracking
- Collaborative analysis
- Communication of the context
- Tracking uncertainty, validity of the analysis, etc.