The term **Provenance** is referred to as ‘The beginning of something’s existence; something’s origin’ and ‘A record of ownership of a work or an antiques, used as a guide to authenticity or quality’[1].

Provenance Tracking is crucial in scientific studies where workflows have emerged as an exemplar approach to mechanize data-intensive analyses. Therefore, research is needed in the area of provenance. However, the current state of provenance with CWL workflows is limited due to various reasons such as:

- **Freedom to utilize any vocabulary for experiment annotation such as EDAM, PROV, FOF**, hence promotes a **domain neutral** standard which can be discipline specific if required by plugging in ontologies explicit for a certain domain.
- **Checksums** for output of each step as well as intermediate data products are saved if specified.
- **Change of parameter space** and **subtraction of a process** for workflow run details and external URIs if any.
- **Affirmation details** annotated by using unique identifiers to establish credit and attribution for researchers to use vocabularies to facilitate the understanding of the results. The complete capture of provenance information will mitigate workflow decay [3] and fulfill applications of provenances to make an experiment transparent, reproducible and authentic. The resulting RO can be distributed using online platform [ROHUB](http://www.rohub.org) which further has quality checks to ensure validity of an RO.

**CURRENT STATE OF PROVENANCE WITH CWL WORKFLOW:**

- CWL provides a standard to write interoperable and portable workflows, therefore an interoperable and portable solution for provenance is proposed in this section.
- Declaring resources as ‘requirements’ at various levels of processing.
- ‘twitfoo’ in input object pointing to the specification hence linking the set of parameters used in a workflow run to the specification.

**RESEARCH PROPOSAL:**

- Aggregates and organizes the workflow outputs, workflow definition, inputs used for workflow run, description of execution environment and attribution details.
- Provides JSON-LD format manifest to describe the research object, listing inputs, outputs, workflow run details and external URIs if any.
- Abstract representation of workflow specification is described using wdfesdc.
- Attribution details annotated by using unique identifiers to establish credit and attribution for publication to facilitate the re-use of execution data.
- Domain-neutral hence flexible in allowing researchers to use vocabularies to facilitate the representation of domain-specific information including title, hypothesis and conclusion.
- Records evolution of a research object by utilizing ‘trees’ ontology.

**REFERENCES**


**CONTACT INFORMATION**

For any Suggestions/Questions Please contact: Farah Zaib Khan (farah.khan@unimelb.edu.au) PhD Researcher University of Melbourne, Australia