Software and Data Provenance as a Basis for eScience Workflow

Joseph Conquest and Michael Stiber
Computing and Software Systems Division
School of Science, Technology, Engineering, and Mathematics)
University of Washington Bothell
Bothell, WA, USA 98011
josephconquest@gmail.com, stiber@uw.edu

***Abstract*—While tracking data and software provenance is necessary in eScience, and often implemented in scientific workflow management tools, such tools generally don't provide graphical UIs that use provenance to guide workflow. Graphitti Workbench was built to center workflow around its PROV-DM based ProVis software and data provenance visualizer, making provenance more readily accessible to scientists. Using ProVis as the system desktop highlights the importance of provenance to inform future experiments and improves provenance understandability. Our usability study indicates that this solution is highly acceptable and that it provides easy deployment of experiments using provenance-based creation.**

I. Poster Summary

This poster presents work in progress towards creating a novel Scientific Workflow Management System (SWfMS) for simulation-driven eScience that centers its user interface around both data *and software* provenance. We outline the major design features of this *Graphitti Workbench*, compare its major features to other SWfMS systems, and present an initial user study.

Here, provenance establishes the relationships among the artifacts associated with a given set of simulations, including executed software, inter-relationships among different software versions, simulation parameters, as well as input and output data. In simulation-driven eScience, the changing behavior of a simulation due to changes in its code poses a risk as software bugs, or other changes, can result in erroneous output. Providing data and software provenance enables scientists to not only determine precise experimental conditions for previous simulations, thereby increasing confidence in results, but also to detect when outputs may be erroneous due to bugs. Data providence alone is not sufficient to make eScience experiments reproducible, as provenance depends on the workflow used for experimentation as well as software provenance [1]. Despite the interconnection between provenance and workflow, few implementations of SWfMS center their design around provenance, instead electing to provide diverse and often incompatible solutions.

Graphitti Workbench, which we will refer to simply as "Workbench", has been developed to work in tandem with the Graphitti simulator [2], [3] to provide an interface for preparing and executing simulations with easy access to information from previous experimentation in the form of provenance [4].

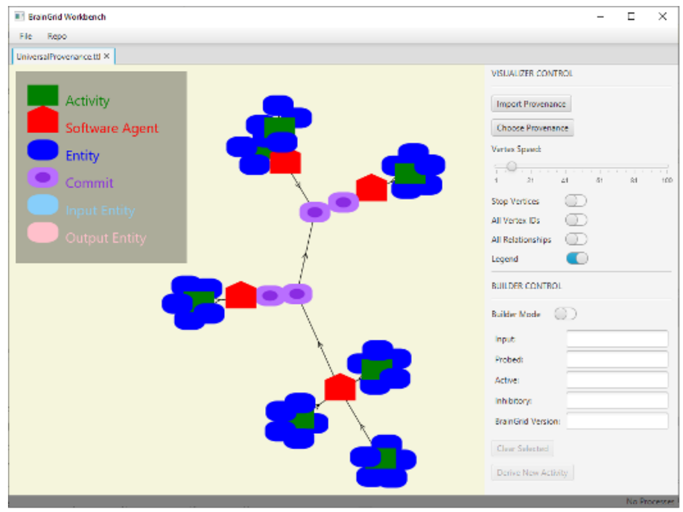

Fig. 1. Workbench ProVis. Legend is displayed at upper left corner and controls are shown to right. Seven different experiments using four different builds of the simulator are shown in the ProVis net.

Workbench provides a graphical provenance visualization tool, ProVis, shown in Fig. 1, which enables users to explore the network of their previous experiments. ProVis uses the PROV-DM standard for visualization, illustrating relationships among agents, entities, and activities as described by the PROV-O standard [5].

Workbench's data and software provenance system addresses the need for comparing different simulator commits as well as input and resulting output [6]. Its ProVis GUI enables users to interact with, and facilitates understanding the interconnection of, data and software provenance. ProVis enables a visual connection between defects found for particular software releases and the likely erroneous data it produced.

A. Comparison to Existing Systems

Currently available workflow systems for eScience provide access to data provenance, but do not provide a holistic view of data provenance's relation to software provenance, nor do they use provenance as a key feature in the development of workflows. They also don't use provenance as a direct input for further research or validation. For brevity's sake, here are three recent examples:

- ReCAP [7] is built on top of Pegasus to produce a graphical visual model similar to PROV-DM. This provenance graph functionality, however, is limited as it cannot be used to reproduce experiments or to compare the provenance of different experiments.
- DQProv Explorer [8] captures and visualizes provenance from data wrangling operations, and provides three visualization components: (i) an explorable provenance graph, (ii) display of quality over time, and (iii) the distribution of issues across entire data sets. The first of these is analogous to the PROV-DM standard. Similar to ReCAP, DQProv's graphical representation is limited in functionality, being unable to reproduce experiments via user interaction.
- Duan *et al.* [9] identify the possibility of using provenance to provide recommendations for future work. They propose using a graph-based uniform workflow provenance model that links design-time and run-time provenance by combining retrospective provenance and prospective provenance. Their implemented solution uses a SQL-like query language to provide this functionality, rather than a visual model.

B. Usability Evaluation

A usability study and survey were conducted in which participants were asked to complete five tasks involving creation and execution of simulations and understanding interrelationships among multiple simulations. Due to the COVID-19 pandemic, only a small number of participants were recruited and evaluation was conducted remotely by a proctor using Zoom. After completing the study, the participant was provided a link to a usability survey for completion in private. For brevity's sake, study methodology details are elided from this poster summary.

User feedback and quality metrics were used to assess user experience and software usability. As software usability is not an absolute metric, we defined it in the sense of who its intended users are, the tasks those users will perform, and the characteristics of the physical, organizational and social environment in which it is used [10]. Given eScience is a domain of both natural scientists and computer scientists, participants for the usability assessment were selected given they had experience with either of these fields. This initial evaluation of Workbench had five participants. The System Usability Scale (SUS) extended by Bangor *et al.* [11] was used as the basis for user feedback, providing an extension of the 0–100 SUS scores to an adjective-anchored Likert usability scale. In addition, time taken to complete the tasks in the study was measured and collected (elided from this poster summary).

Table I shows the calculated SUS scores and average for each participant, as well as Bangor *et al.*'s overall quality metric. Bangor's metric is scaled 1–7, with 1 representing "worst imaginable," 2 "awful," 3 "poor," 4 "ok," 5 "good," 6 "excellent," and 7 "best imaginable." The average of Bangor's overall quality metric was 5.8 out of 7, or a quality percentage of 82.86%, corresponding most closely to the "excellent" descriptor for the usability of Workbench. Combined with the average SUS score, 74.5, this shows Workbench usability falling between the "good" and "excellent" adjective rating, as shown in Fig. 2.

TABLE I
PARTICIPANT RESPONSES TO USABILITY

Evaluation Metric	**Participant**					
	1	***2***	***3***	***4***	***5***	***Average***
Calculated SUS Score	65	77.5	85	82.5	62.5	74.5
Bangor Overall Quality Metric	5	6	6	6	6	5.8

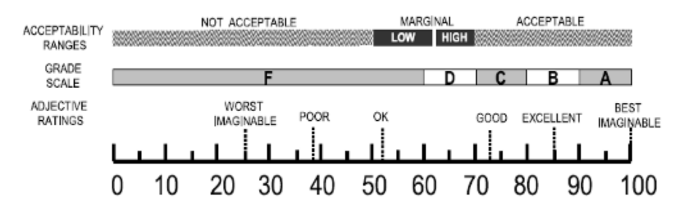

Fig. 2. The adjective ratings, acceptability scores, and school grading scales in relation to the average SUS score provided by Bangor *et al.* [11].

REFERENCES

[1] L. Pouchard, S. Baldwin, T. Elsethagen, S. Jha, B. Raju, E. Stephan, L. Tang, and K.K. Van Dam, "Computational reproducibility of scientific workflows at extreme scales", *Int. J. High Perf. Comp. Appl.*, vol. 33, pp. 763–777, 2019.

[2] C. O'Keefe, T. Salvatore, M. Stiber, K. Dukart, and L. Presland, UWB_Biocomputing/Graphitti: Dangerous Scribbles. Zenodo, 2021 [Online]. Available: https://zenodo.org/record/4678632/

[3] Stiber, M., Kawasaki, F., Davis, D. B., Asuncion, H. U., Lee, J. Y., and Boyer, D., "BrainGrid+Workbench: High-Performance/High-Quality Neural Simulations", *2017 Int. Joint Conf. Neural Networks (IJCNN)*, Anchorage, AK, 2017, pp. 2469-2476, doi: 10.1109/IJCNN.2017.7966156

[4] Conquest, J., Vetrivel, K., Stiber, M., Presland, L., Wright, M., and Shastri, H., UWB-Biocomputing/Workbench: Conquering Hero. Zenodo, 2021 [Online]. Available: https://zenodo.org/record/4678834/

[5] K. Belhajame, J. Cheney, D. Corsar, D. Garijo, S. Soiland-Reyes, S. Zednik and J. Zhao, "PROV-O: The PROV Ontology", World Wide Web Consortium (W3C), 30 April 2013. Available: https://www.w3.org/TR/prov-o/. [Accessed 17 November 2020].

[6] D. Davis, H. Asuncion, and M. Stiber, "Software and Data Provenance in Computational Neuroscience Research", *11th IEEE Int. Conf. eScience*, Munich, Germany. 2015.

[7] K. Hasham and K. Munir, "Reproducibility of scientific workflows execution using cloud-aware provenance (ReCAP)", *Computing*, vol. 100, pp. 1299–1333, 2018.

[8] C. Bors, T. Gschwandtner, and S. Miksch, "Capturing and Visualizing Provenance from Data Wrangling", *IEEE Comp. Graphics and Applic.*, vol. 39, no. 6, pp. 61–75, 2019.

[9] X. Duan, J. Zhang, Q. Bao, R. Ramachandran, T.J. Lee, S. Lee, and L. Pan, "Linking Design-Time and Run-Time: A graph-based Uniform Workflow Provenance Model", *IEEE 24th Conf. Web Svcs.*, 2017.

[10] J. Brooke, "SUS: A 'quick and dirty' usability scale", *Usability Evaluation in Industry*, Taylor & Francis Ltd, 1996.

[11] A. Bangor, P. Kortum, and J. Miller, "Determining what individual SUS scores mean: Adding an adjective rating scale", *J. Usability Studies*, vol. 4, issue 3, pp. 114–123, 2009.