

The PBase Scientific Workflow Provenance Repository

Victor Cuevas-Vicenttín, Parisa Kianmajd, Bertram Ludäscher, Paolo Missier, Fernando Chirigati, Yaxing Wei, David Koop, Saumen Dey

International Journal of Digital Curation, ISSN 1746-8256, Vol. 9, n. 2, 2014, p. 28-38

Los flujos de trabajo científico y sus sistemas de apoyo son cada vez más populares para los experimentos científicos de cálculo y uso intensivo de datos. Las ventajas que los flujos de trabajo científico ofrecen incluyen el diseño de un flujo de trabajo rápido y fácil, reutilización de software y datos, ejecución ampliable, intercambio y colaboración, y otras ventajas que facilitan en conjunto “ciencia reproducible”. En este contexto, la procedencia – información sobre el origen, contexto, derivación, propiedad, o historia de algún artefacto – juega un papel clave, ya que los científicos están interesados en examinar y auditar los resultados de experimentos científicos. Sin embargo, para realizar tales análisis sobre resultados científicos como parte de colaboraciones de investigación ampliadas, se requiere un entorno y herramientas adecuadas. En concreto, la necesidad surge de un repositorio que facilite el intercambio de flujos de trabajo científico y sus trazas de ejecución asociadas de manera interoperable, permitiendo también la consulta y visualización. Además, esta funcionalidad se debe apoyar mientras tenga en cuenta rendimiento y escalabilidad. Con este propósito en mente, introducimos Pbase: un repositorio original de flujo de trabajo científico que lleva a cabo la propuesta ProvONE, la cual amplía el estándar emergente W3C PROV para datos de procedencia con conceptos específicos de flujo de trabajo. Pbase se construye sobre la base de datos gráfica Neo4j, que ofrece capacidades como la consulta declarativa y eficiente. Nuestras experiencias demuestran el poder adquirido mediante el apoyo de diferentes tipos de consultas para los datos en origen. Además, Pbase está equipado con un interfaz ágil para el usuario y a medida para la visualización de flujos de trabajo científico de datos en origen, haciendo más fácil y eficaz la especificación de consultas e interpretación de sus resultados.

Resumen traducido de la propia publicación