Velo: A Collaborative Data Management, Analysis, Simulation, and Publishing Environment

Velo is a reusable, scalable, domain-independent infrastructure for managing scientific work. It supports the entire scientific project life cycle, including data management, modeling and simulation, visualization and analysis, validation, reporting, archival, publishing, and discovery. Velo is built on extensible open-source technologies to create a collaborative core platform that can be tailored to specific scientific uses and deployed to new sites in weeks. Developed over the past 10 years and in production use by more than 15 separate projects and services supported by the U.S. Department of Energy (DOE) and other agencies, some of Velo's unique features include:

Scientific Data Management

Velo extends standard content management features (versioning, metadata capture, access control, indexing, etc.) to support scientific work. This includes a customizable scientific content model for describing projects, data sets, tools, and investigators as well as provenance capture (i.e., relationships between data, tools, and people to form an "audit trail"), stored in a graph-based knowledge store. In addition to automated provenance capture, Velo provides a sophisticated markup capability where users can add annotations to data through a Wiki interface. For more detailed provenance capture and to meet reproducibility requirements, Velo can be extended with the PNNL ProvEn provenance environment. Velo also allows users to integrate, reference, and access not only previously uploaded data, but geographically distributed heterogeneous data as well.

Tool Integration

Velo provides a rich infrastructure, including a tool registry and rich client interface for integrating third-party applications. Tools can range from desktop applications to workflow management systems and high-performance simulation codes. This approach enables Velo user communities to leverage their existing application portfolio to create an efficient collaborative science environment that seamlessly combines simulation, experiment, data, and analysis with repeatable workflows and provenance capture.

Simulation Management

As a platform for high-end simulation management, Velo provides a job execution infrastructure that can interact with any number of remote high-performance computing platforms and scheduling systems. The infrastructure will handle job setup and submission and monitor the job execution, including error reporting and mitigation. Finally, the output will be returned and registered with Velo, and a provenance record is created for the job execution. Velo can offer additional support to fully manage large ensemble, combinatorial, or uncertainty quantification runs, overseeing the complex setup, staging, submission, and monitoring processes. Velo also can interact with most workflow management systems and provide a simple interface for executing and monitoring complex workflows.

Data Publication

Velo provides a data publication service where data sets are published with a unique Digital Object Identifier (DOI) and registered with DataCite. This data publication service ensures long-term preservation and supports better linkage between DOE's published research results and the underlying data. The DOI for all registered data sets provides access to an introductory page (landing page), as well as the data set and any related publications (access permitting). Existing Velo users will have full access to the published data and tools to interact with the data and supplementary information, such as presentation, white papers, and publications.

Highly Customizable User Interfaces

The Velo platform offers two primary user interface layers—a machine-independent rich client interface and a web-based user environment. Each user interface layer is highly customizable and allows the creation of easy-to-use, tailored interfaces for different domains, so users can manage projects, applications, and data sources in intuitive ways. Clear application program interfaces for accessing the Velo server functionalities also afford integration with other user interface environments.

Security

Velo, which sits atop the Alfresco ECM platform, includes an extensible, fine-grained, role-based access control framework based upon Spring Security (*http://projects.spring.io/spring-security/*) that can be configured to work with a number of authentication providers, including LDAP, Active Directory, NTLM, Kerberos, OpenID, Shibboleth, and OAuth. Configuring Velo to use a federated identity provider, such as OAuth, enables Velo to communicate seamlessly with external tools, data, and compute resources in a single-sign-on (SSO) environment. The Velo login screen can be configured to support various credentials, including username/password; X509 certificates; and multi-factor authentication (MFA), such as pin and SecureID token. Velo access controls can be applied using an intuitive interface for managing users and teams and setting user- and team-based permissions on resources, such as projects, data sets, or files.

PNNL Velo Team

Carina Lansing, Zoë Guillen, Gary Black, Chandrika Sivaramakrishnan, Sumit Purohit, Ellen Porter, and Bibi Raju
velo-team@lyris.pnnl.gov

Business Contact

John R. Johnson
john.johnson@pnnl.gov

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF
ENERGY

Proudly Operated by **Battelle** *Since 1965*