

Abstract

Pacific Northwest National Laboratory Resource Discovery for Extreme Scale Collaboration

Karen L. Schuchardt, Lead Principal Investigator, 509-375-6525; karen.schuchardt@pnnl.gov
Peter Fox, Ph.D., Co-PI, Rensselaer Polytechnic Institute

At the core of any collaborative infrastructure is the ability to discover, manage and utilize relevant resources accessible to the collaboration. These include not only computing, experimental and storage systems, but also people, data, applications, and services. Equally important are the relationships between resources, which capture critical contextual knowledge that enable their effective use. Resource description databases have been part of collaboratories, software systems and web applications because they fill a crucial role in driving user level tools. However, existing capabilities are typically designed to fulfill a limited set of requirements for a particular collaboratory, support a targeted and limited set of resources, and do not support ad-hoc resource descriptions and linking and the necessary accompanying discovery mechanisms.

Our objective is to develop a capability for describing, linking, searching and discovering resources used in collaborative science that is lightweight enough to be used as a component in any software system such as desktop user environments or dashboards but also scalable to millions of resources. A key design goal is to offer local control over resource descriptions thus reducing one of the bottlenecks to widespread adoption.

We propose to build a prototype framework and associated services, the Resource Discovery for Extreme Scale Collaboration (RDESC), that meet these objectives. Though the PIs have extensive experience building user environments upon which to draw, we will first develop a set of use-cases across several science domains to serve as the driver for our design and eventual implementation and demonstrations. For the purpose of capturing semantic of context, we will adopt sets of existing ontologies where possible. Key contributions of our work will include: several means of publishing information to be both relevant to the type of resource and to create a low barrier to entry; a variety of browse, search, and discovery interfaces; and scalable hybrid data stores. Our approach will be to leverage existing standard Web languages and protocols, and open source tools and libraries where possible.

The outcome of this work will be documented use-cases, a testbed that evaluates one or more solutions for scalable discovery, a software toolkit that can be incorporated into other systems, and a demonstration of RDESC discovery interfaces. If successful, we will have created a means to describe critical scientific resources as a global, discoverable set of resources and eliminate the isolated islands of resource description that exist today. We believe that such a system can greatly increase collaboration and new research directions because an important current bottleneck is caused by lack of awareness of what is available and continued reliance on personal information exchange