**Research SaaS as a basis for high-performance collaboration**

Comments to Ian Foster <foster@anl.gov>

The distributed and collaborative nature of modern research means that a typical investigator is constantly involved in a dynamic mix of resource sharing and teaming relationships of widely varying purpose, scale, and scope. For example, a computational researcher can easily during a single morning need to access code in a source repository shared with a code team, run code on a supercomputer at another lab, share results with a multi-lab science team, engage with experimentalists to obtain data for validation, and share the results of a validation study with yet another group, all the while recording the steps followed in this study for future reference by the entire community.

**Obstacles to collaboration**: Despite considerable advances in such areas as networking, security, and data movement (three areas in which DOE innovations have been influential), the establishment and secure and efficient operation of such resource sharing and teaming relationships remains inordinately difficult. It is not uncommon for the apparently simple act of sharing a single dataset to involve many email exchanges and manual manipulation of accounts, permissions, and privileges—with results that are frequently inefficient, nonintuitive, and even insecure.

A major reason for such difficulties is the continuing bespoke nature of much of the technology used to enable collaboration. Requiring researchers to install software, manage research data, configure collaboration tools, or manage permissions is like requiring them to install and run their own telephone exchange to make a phone call—a telephone exchange that will not interoperate with those installed by other teams.

**Research SaaS**: We argue that the answer to this problem is to get individual researchers, research teams, and laboratories out of the business of installing and operating the software and other information technology (IT) used in their research. Instead, responsibility for much of that IT should be transferred to software as a service (SaaS) providers. In industry, this approach has been tremendously successful: many small and medium businesses today obtain essentially all of their IT functions (payroll, accounting, Web presence, email, customer relationship management, etc.) from business SaaS providers. Equivalent services for research would allow an individual or team (a *small or medium research laboratory*) to outsource to research SaaS providers the task of establishing and operating the collaboration, data management, security, and other tools required for an effective resource sharing or teaming relationship. Such providers would provide on-demand access to powerful *virtual laboratories* that individually or collectively deliver all of the IT required for effective research.

**An illustrative scenario**: The researcher mentioned in the first paragraph would, in this approach, make use of multiple virtual laboratories during their morning’s work: one for collaborative source code development (think Github for HPC, with appropriate collaboration, testing, test coverage, and related tools); another for supercomputer access, providing for example task management functions; and yet another for managing the data products and data analysis tasks associated with an experimental campaign. Underpinning these various collaborative spaces would be common credential management, group management, audit, and provenance services, to reduce barriers to resource sharing and teaming, and provide for reproducible research.

**Potential advantages**: Experience with SaaS for small and medium businesses and in the consumer market (e.g., Google collaboration tools, Netflix, Apple services) suggest that outsourced research IT services may be able to deliver dramatically improved capabilities, ease of use, and economics relative to software that is installed and operated on local computers. Modern Web 2.0 technologies can provide intuitive interfaces to powerful capabilities, reducing barriers to their use. The operation of these services by specialist providers can improve quality of service and achieve economies of scale. Centralization can also facilitate information sharing and collaboration (compare and contrast how photo sharing works when photos are on your home hard drive vs. uploaded to Flickr) and provide a context for deployment of value-added services, such as automated metadata generation and data curation functions. Whether comparable benefits can be achieved in the research space remains to be demonstrated, but early experiences with research SaaS systems such as Globus Online and nanoHUB have been extremely encouraging.

**Research challenges:** The successful realization of this vision of outsourced IT will require answers to a range of challenging research questions. What are the critical processes that underpin modern research—the equivalents for small and medium research teams of payroll, accounting, and CRM for small and medium businesses? What are the foundational elements on which can be built robust, secure, and scalable research data management and collaboration solutions? How can these elements be integrated with campus and national cyberinfrastructure systems, such as supercomputer centers? How do we scale solutions to massive data, large teams, and high-throughput processes? How do we integrate the audit and provenance mechanisms required for reproducible research, without creating onerous requirements on investigators? What user experience (UX) elements are important in research? (Companies such as Netflix, Google, Apple, and Amazon have pioneered approaches to consumer UX that have proved transformative in their usability. Will similar methods work for science?) How does outsourced IT change the security equation? (SaaS can enhance security by professionalizing its operations. But presumably it also introduce new risks. How can those risks be managed?) It will also be important to study the economics and sociology of DOE science. What factors may hinder or encourage adoption? How will we sustain such outsourced services? What may be the unexpected consequences?