Project Sushi
Enterprise Data Management for the EH&S Chemical Suite of Applications

submitted by

Jay Ballinger
UC Systems Architect
530-554-1373
jayballinger@ucdavis.edu

project team

Jay Ballinger
Diana Cox
Safa Hussain
Mira Kaloper
Serguei Mysko
Ed Storm
Stefan Tomic

There has been increased focus and scrutiny on our University of California research laboratories – particularly those laboratories which store and use dangerous chemicals. The UC Office of the President (UCOP) responded to the need for more safety and attention by sponsoring a suite of applications concerned with the safe identification, accounting, usage, and disposal of these chemicals. Additionally, emphasis was placed on identifying safe handling procedures and proper protection equipment which will help to keep our researchers safe. These applications — collectively called the UCOP ERM Chemical Suite — are designed to specifically address the many aspects of acquiring, storing, using, and disposing of these chemicals.

During the envisioning and design of these six applications, it became apparent that they share many of the same requirements. At their core these applications are concerned with identifying the Principal Investigators and their Lab Workers, the locations of their laboratories, and the storage and use of chemicals used in their research. These three categories of data — people, places, and chemicals — presented a significant challenge. Never before has this data been collected across all UC campuses in any organized fashion. The collection of this data required unprecedented cooperation from multiple departments on each campus.

To provide an example of the effort consider just one of the Chemical Suite applications – the Chemical Inventory System (CIS). CIS requires each Principal Investigator (PI) to identify the locations where chemicals are stored and used. The CIS application uses this data to provide an accounting and to also produce a safety compliance report for the campus to be shared with multiple safety teams. To do this
effectively we need to start with a list of PIs and their Lab Workers (LWs) who might be delegated the task of accomplishing the inventory. Each campus has its own method for identifying people and their roles on campus, but the relationship of PI to LW was very difficult to accumulate. Even the definition of PI is not consistent across every context. The identification problem is compounded by a lack of a truly globally-unique identifier. UCNetIDs are not issued to everyone, so a combination of UCNetIDs, ePPNs, and other ‘best available identifiers’ were used to help identify people and then we used user-supplied relationships to build the associations between them.

The next challenge is the identification of laboratory locations. The best source of building and room data is each campus’ facilities system of record. Luckily many campuses use a common application platform for managing their physical locations, but this common platform was not suited well for multiple on-demand ad-hoc queries to find buildings and rooms by common names used by the researchers. And for the campuses that used a custom solution, providing multiple applications access to their database is not considered a technological best practice.

Finally, access to a list of chemicals from which the PI or LW could choose and then identify the quantity stored and used within their laboratory was needed. Chemicals can be identified by multiple means – using names, synonyms, registry numbers, etc. – and their physical properties are used to help determine their potential for danger. The best source for chemical data comes from the chemical industry via the manufacturer and other sources that provide descriptions and useful classifications.

These are heavy-weight requirements for any single application. Acquiring the people, relationship, and location data from each campus is a daunting task. Combine that with the technological need to store, update, and report on this data and one can imagine the development effort needed to accomplish this. Add to that the very large list of chemicals that can be used on a campus, and how to acquire and manage that data as well, and the workload on the development team gets bigger.

Now, take these large requirements needed just to support and application like CIS, and multiply the work load by six. Each of the applications in the Chemical Suite has very similar needs in their combination of people, places, and chemicals. It was quickly determined that if a central coordination of this data was not created, then each of these applications would be engaging the campuses and the industry one after the other. It would be easy to visualize each application project team independently asking the campuses for their list of PIs, their list of buildings and rooms over and over for each project. Certainly the campuses would begin to question why the same data is being requested by multiple teams multiple times over the course of many months.

Project Sushi was created to bring consistency and Enterprise Data Management to support not only these Chemical Suite applications, but to future UCOP centralized applications as well. Sushi will be the
centralized publishing point of research personnel and their roles, buildings and rooms on each campus, and chemicals and their descriptions. The data will be submitted or gleaned from the various campus systems of record and from industry partners and will be made available to client applications – like the Chemical Suite – via web services for use within those applications. Additionally, Business Intelligence reporting is now possible because of the consistent use of people, places, and chemical identifiers shared across the client applications. Sushi will provide one-stop-shopping for this essential data saving time, effort, and money by allowing the client applications to focus on their unique functionality and not on redundant data acquisition and storage.

Technology Used

Project Sushi derived its name from the technology being used. Essentially Sushi needs to manage data in the following domains: campus, people, roles, buildings, chemicals. Each of these domains needs to be supported by a database, needs to provide a user interface for basic management of the data (inserting, updating, and deleting), needs to publish this data via web services, and needs to provide proper security – or authorization – in allowing client applications to provide new data or to query the existing data.

The project team was faced with the classic “Build versus Buy” decision. Do we build such an application infrastructure framework from scratch, or did something already exist which provides many, if not all, of the basic technological pieces? Building a new system from scratch is always tempting because of the challenge and the ability to build exactly what is envisioned. But building from scratch usually takes a significant amount of time – time that could be spent on the custom functionality needed rather than the underlying platform. For Project Sushi the Kuali Rice middleware framework was chosen to be the basis for our custom application development. Through that platform we were able to model and accommodate the data structures needed, leverage the authentication and authorization functionality, employ the workflow features, and publish the business logic on a service bus.

It is a unique use of Rice to use it as a platform for custom application development, but it was the logical choice given the requirements, the in-house experience, and the technology stack compatibility with our desired programming language. It saved us time and money by allowing us to focus on the custom business logic needed to coordinate the data, and not focus on the mundane task of data persistence and publishing. Our custom code is the good stuff sitting on top of Rice – kind of like a nice piece of sushi. Thus a project name was born and many sushi lunches consumed.
The Value and Savings of Project Sushi

Unlike a mobile application or website, Project Sushi is not used directly by the end-users. Rather it is part of the magic behind the scenes that provides the heavy-lifting freeing the client applications to focus on their specific functionality. In other words, it provides an economy of scale by not requiring each application to duplicate data and effort. The design, work, and data-gathering efforts can be done once and shared across the client-application suite as a centralized repository.

The investment in Sushi pays dividends when each new client application subscribes to its data. The reuse of Sushi’s data and functionality not only provides common functionality across the application suite, it provides data consistency as well. Each client application has access to the same data, the same unique identifiers, and contributes to the value of the centralized data by presenting the same consistent data to the end users. The PIs and LWs will recognize the same laboratory room definitions across all of their applications. They will define their PI to LW relationship just once and have that relationship shared across the applications. The users will enjoy a single comprehensive list of chemicals and chemical attributes and will reliably identify these chemicals across all of the applications.

Development and Implementation Timeframe

The timeframe for development and implementation was very short. By the time we realized the need for Sushi to centrally publish the people, places, and chemical data, we were already behind schedule. The other applications of the Chemical Suite had some immediate needs and some aggressive delivery dates. Sushi was envisioned, planned, developed, and deployed in just six months. It will continue to grow in depth and in features as new opportunities are identified.