Significance of the EdUnify Project

The web service registry approach to collaboration and reuse of web services has worked effectively in other endeavors. A web service registry is at the heart of the cancer Biomedical Informatics Grid® (caBIG®), an information network funded largely by the National Cancer Institute, enabling members of the cancer community to share data and knowledge.[1] The technical infrastructure of caBIG® is a ¿grid¿ of compatible web services enabled by a web service registry and core interoperability services known as caGrid. The caGrid portal provides a view into the registry of services, which at the time of this writing contains 129 web services from 124 participating institutions.[2] The caGrid Index Service or registry allows members of the cancer community---researchers, physicians, and patients---to share information and knowledge. For example, the grid allows physicians to locate and reuse complex analytical resources on the web, it enables researchers to query large data sets at multiple institutions in a single query, and it manages patient study calendars for clinical trials. The caGrid registry infrastructure provides the core foundation for innovation. Members of the cancer community can find what programmatic resources are available on the web, where they are located, and how to invoke these resources. This transparency fosters the invention of new web services based on further analysis and research, the generation of new hypotheses, and the execution of research which accelerates the discovery of new strategies for the detection, diagnosis, treatment, and prevention of cancer.

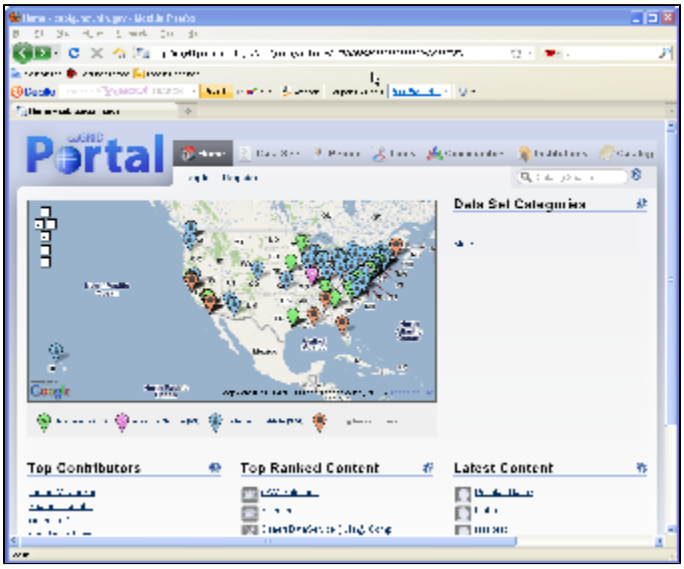

Figure 1: caGrid Portal (click to enlarge image) displaying a visual representation of the web services in its Index Service or registry. The data web services, analytical web services, and participating institutions appear as pins on the map. Drilling down on the map provides details of the services and how to invoke them. The caGrid portal also aggregates data about the contributors and quality of the web services.

Another example of the successful application of web service registry technology is the BioCatalogue, a registry of biological web services funded by the Biotechnology and Biological Sciences Research Council (BBSRC) in the United Kingdom.[3] At the time of this writing, the BioCatalogue has 1,694 services registered from 124 service providers. The BioCatalogue, while available to the public without registering for a user account, also has 397 registered users. These core users actively register web services and provide valuable data about the services (metadata), they annotate web service entries with details about how to invoke the services, provide alternate names of data fields to help map equivalencies for interoperability, and rate the value of web services. This user community, supported by the social collaboration features of BioCatalogue, fuels the discussion about re-use, the development of new services, and inter-institutional research collaboration. The BioCatalogue software also monitors the services in the registry and reports on their availability. This metadata helps potential users of the web services determine which services are best to use and has led to the development of more advanced interoperability and orchestration services that can make use of this metadata. For example, semantic web tools can use this metadata to perform automatic ¿reasoning¿ about which data fields may be equivalent in data sources on the web. This reasoning allows the author of a report to include many more data sources in the report than otherwise possible. If the report writer had to perform manual analysis of all data sources alone, without the help of the user community providing these annotations, a report of the same scope would take much longer to prepare and, in some cases, the time and cost would be prohibitive.

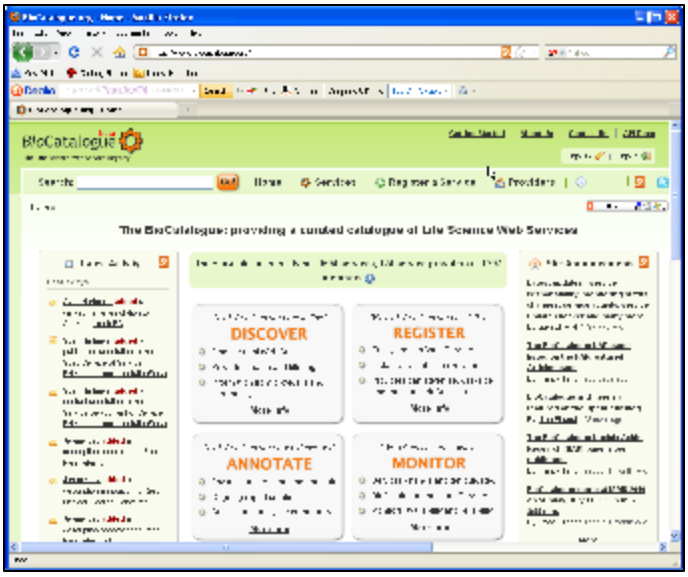

Figure 2: the BioCatalogue registry (click to enlarge image) for discovering, registering, annotating, and monitoring web services for the life sciences listing 1,694 services from 124 service providers. This view shows the latest activity of users registering and annotating services along with announcements about latest integration tools for web services that leverage the registry like Taverna Workbench.

In both the case of caGrid and BioCatalogue, the existence of the web service registry led to the development not only of new services enabled by the existing services in the registry, but also more advanced tools to help developers compose new services and integrations more quickly and effectively. For example, Taverna Workbench is an open source development environment that allows users to import web service definitions into and build integrations or workflows that invoke these services.[4] Taverna is useful for developing integrations or reports that query any services developed using web services technologies. However, its true potential is achieved when it is combined with a web service registry as with caGrid, with BioCatalogue, or in the future with the EdUnify registry. An integration developer can point the Taverna Workbench to such a registry and immediately import all of the web service definitions into the workbench. Then the developer can draw from the palette of data services to build a query or workflow to implement the task at hand. The integration analyst or developer does not have to perform extensive research about which services are available from each data source. Not only is this approach more efficient than coding these integrations with lower-level tools, but it also encourages re-use of common data standards and existing integrations by directly presenting the integration analysts and developers with existing service definitions and implementations.

There are a number of free and commercial integration development tools that leverage web services technologies and registries like Taverna Workbench. However, Taverna is especially interesting because it demonstrates how web service technology standards and web service registries lay the groundwork for further innovation. In addition to making the development of web service invocations easier, Taverna also allows users to save and share the integrations they develop to another registry of workflows, called myExperiment, dedicated to these integration orchestrations. So, in many cases, developers of these web services integrations can visit myExperiment and find a workflow that already meets their needs or a workflow that can be tweaked quickly for their specific use case.[5]

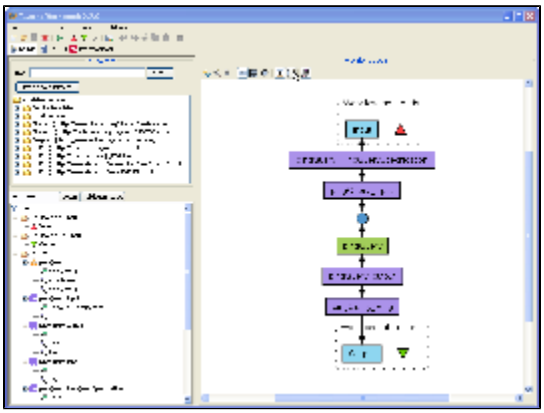

Figure 3: the Taverna Workbench (click to enlarge image) graphical development environment for modeling and executing workflows with web services. This view shows a flow designed to invoke the OpenEAI Academic History Web Service ping operation. This web service is presently deployed at Tri-County Technical College, Trident Tech, Midlands Tech, and the University of South Carolina as part of South Carolina¿s SCTrac.org initiative to facilitate transfer of credit between institutions.[6] All of these services can be found in the EdUnify Demo Registry.

The EdUnify project idea was conceived by members of PESC as an application of the web service registry concept that has been used successfully in other contexts such as cancer research and life sciences. In 2009 the members of PESC discussed the far-reaching benefits of a web service registry for higher education.

- If online course transfer and articulation service providers could discover and invoke web services at institutions that provide specific transfer disclosure information by course with transfer policies, these service providers could develop web applications to help students model transfer scenarios more quickly and completely than they can today.
- If online service providers could discover and invoke web services for course catalogues at academic institutions, they could develop web applications to deliver more complete and current data to inform the course articulation decision process. These applications could also guide students to courses that will transfer to their desired target institutions without having to rely on paper copies of course catalogues and extensive searches of the web.
- If online course transfer and degree advising service providers could discover and invoke web services at institutions to import students¿ academic histories of coursework, these service providers could more advise students more quickly and accurately. Although online course transfer and advising services exist today, they are limited in their effectiveness because they rely on students to provide their own academic histories by manually typing them in, which is time consuming and prone to error. Automating the academic history import process will dramatically increase the adoption and accuracy of online course transfer and advising services.
- If government agencies could discover and invoke web services at institutions that fulfilled mandated reporting requirements, potentially more data could be gathered at more frequent reporting intervals to inform policy, while placing less burden on the academic institutions to prepare and deliver these reports themselves.

These ideas and many other potential benefits led the members of PESC to convene the EdUnify Task Force.

The EdUnify Task Force met for the first time in December 2009 in Washington, DC. The initial meeting was attended by 40 representatives from 30 organizations across higher education, representing colleges, universities, system vendors, standards bodies, and government agencies. The group agreed that such a registry would be valuable infrastructure. The group also recognized that such a registry had the potential to promote innovations for higher education the group could not yet conceive---just as the interstate highway system, the internet, and the mobile phone network infrastructure was designed with a specific vision and purpose, but enabled many other innovations beyond their initial scope.

The EdUnify Task Force formed technical and business work groups to focus on different aspects of building a web service registry for higher education. By March 2010, the group had reached several conclusions. First, the technology already exists from the development of other web service registries such as caGrid and the BioCatalogue to demonstrate the basic functionality of a web service registry for higher education. The group resolved to deploy such a demonstration registry with volunteer effort and limited funding from PESC members. The members decided to use the free BioCatalogue software available from that project and use its demonstration deployment to help the EdUnify project participants develop additional requirements that may be specific to higher education. Second, the business team developed a draft business plan to develop and sustain a web service registry project, but concluded there were significant challenges to overcome in the startup phase. Specifically, the team observed that nearly all of their organizations were prevented from making significant contributions by their diminished capacity to perform new work. The group recognized that it would be challenging to invest in improving the infrastructure for integration and interoperability, because their organizations were already overcommitted to existing work. In order to break the cycle of inefficiency and increase capacity with new infrastructure like a web service registry, some external funding would likely be required.

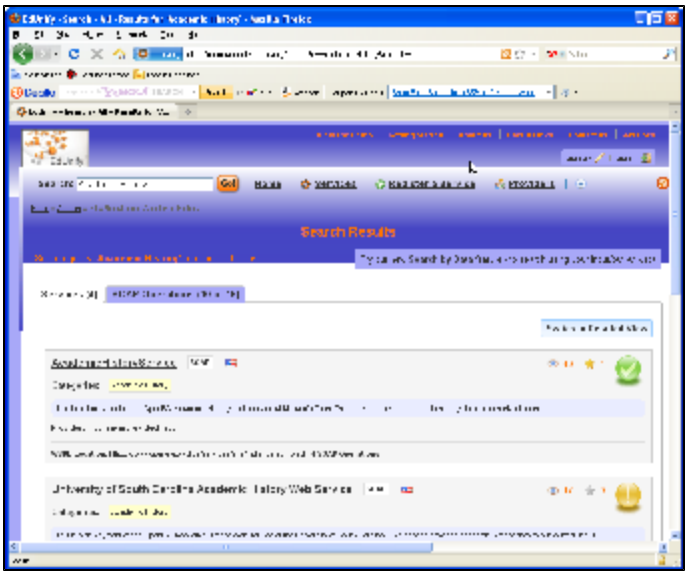

Figure 4: the EdUnify Demo Registry based on the BioCatalogue (click to enlarge image). Displayed in this view are the first results from a search of the registry for ¿Academic History.¿ These services are deployed at Midlands Technical College, the University of South Carolina, Trident Technical College, and Tri-Country Technical College in South Carolina. The EdUnify Demo Registry is available at https://demo.edunify.pesc.org.

Read more about the need for the EdUnify project.

---

[1] National Cancer Institute, ¿About caBIG®,¿ National Cancer Institute, https://cabig.nci.nih.gov/overview.

[2] National Cancer Institute, ¿caGrid Portal,¿ National Cancer Institute, http://cagrid-portal.nci.nih.gov.

[3] Jiten Bhagat, Franck Tanoh, Eric Nzuobontane, Thomas Laurent, Jerzy Orlowski, Marco Roos, Katy Wolstencroft, Sergejs Aleksejevs, Robert Stevens, Steve Pettifer, Rodrigo Lopez, Carole Goble: ¿BioCatalogue: a universal catalogue of web services for the life sciences,¿ *Nucleic Acids Research* (2010). doi:10.1093/nar/gkq394. Online at: http://nar.oxfordjournals.org/cgi/content/full/gkq394.

[4] University of Manchester, ¿Taverna – open source and domain independent workflow system,¿ University of Manchester, UK, http://www.taverna.org.uk.

[5] University of Manchester and the University of Southhampton, ¿myExperiment Virtual Research Environment,¿ http://www.myexperiment.org.

[6] The South Carolina Transfer and Articulation Center (SC TRAC) is a service center for transfer students, administrators, advisors, and faculty, designed to facilitate student transitions between South Carolina¿s institutions of higher education. See http://www.sctrac.org for details.