



> Semantic Web Use Cases and Case Studies

Case Study: POPS — NASA's Expertise Location Service Powered by Semantic Web Technologies

Michael Grove, Senior Software Developer, Clark & Parsia, LLC, and Andrew Schain, The HQ CT/EA, NASA

February 2008



Introduction

NASA collects and stores information regarding an employee's organizational affiliation, work history, evidence of skills, and details to derive with whom the employee has worked. The problem is that this information is fragmented across different databases, application silos, and other systems. If you are searching for an individual with unique skills and work history, you need to query across multiple systems. Even if such queries were easy, most people would still want to ask for a personal reference from someone they trusted.

This "rolodex culture" is not unique to NASA. If you are looking for an expert in an engineering discipline in order to staff a new project, or to form a working group, there are only a couple options for finding individuals. Either you know someone already at NASA who is qualified for the job, or you can ask someone at NASA if *they* know someone who is qualified for the job. When this works, it works very well. When it doesn't work, it's a waste of time at best. In either case, this technique doesn't work at all if you're really interested in workforce planning, management, or analysis; or if you're interested in aggregates, not individuals; or if you need the kind of information that can really only be derived from integrating multiple information sources together.

All of this presents difficulties for anyone trying to staff an upcoming project or trying to form a working group. The staffer must rely primarily on the "social reach" of their personal network of contacts inside the Agency. While such interactions do have social and institutional benefits, those are often not directly germane to "expertise location". Finding a "new" person is extraordinarily difficult; and when you do manage to locate someone, their contact information is not stored alongside their core competencies, where they work, what projects they have participated on, or publications they have authored.

The aim of POPS is to integrate NASA's information about its nearly 70,000 combined civil service and contractor workforce in one place, linking the relevant, related information to form a comprehensive data service for staffers, workforce planners, analysts, and related personnel. POPS delivers an easy to use, effective application for expertise location and workforce analysis in a cost-effective manner by reusing existing information sources, and integrating them using RDF and other Semantic Web technologies. POPS provides access to the resulting information aggregation via *jSpace*, which is a visual query builder and Linked Data browser for SPARQL and other RDF query languages (<http://www.clarkparsia.com/jspace/>).

POPS is a NASA-wide production application and is believed to be NASA's first non-mission, Semantic Web application that is available to all NASA employees.

General Description

Employee information at NASA exists in many different places. Contact information is stored in an LDAP server, project information is stored in a time and attendance accounting system, employee competency information is stored in a human resource system; and all of these are at geographically distinct NASA facilities. Further, the NASA technical reports repository is stored in a public database at yet a different NASA facility. By aggregating all of this information into an RDF database, which is accessible via a lightweight, HTTP-based web service using SPARQL (and SerQL) query languages, application developers, including primarily the POPS team, can leverage the resulting information integration to build domain-specific applications and other services. Thus, using Semantic Web technologies in the non-mission side of NASA IT has meant the creation of a lightweight information integration infrastructure and the POPS expertise location service. Since clients of the origin data sources feed this aggregation, those data sources are left undisturbed and fully operational. The POPS team consciously pursued a policy that did not insist on centralization and mothballing of the constituent data sources, as that approach is far too disruptive and would not provide appropriate ROI.

However, aggregating the data was only the first step, users needed a tool which would help them browse the data without requiring them to learn a query language. To that end, Clark & Parsia LLC developed *jSpace*, a general-purpose visual query builder and Linked Data browser for RDF databases. *jSpace* is a simple, powerful tool which translates user input and user interface state into RDF queries that can be executed by the remote RDF database. *jSpace* was initially developed as a direct Java clone of *mSpace* (<http://mspace.fm>), but soon diverged as many extensions needed to be added to the User Interface (UI) and application model.

One of the key components of *jSpace* is the ability to plug-in different UI widgets that can provide custom views of the data being browsed. Along with the default "Details" view which shows all the property-value pairs of a selected resource, we developed several custom components for POPS.

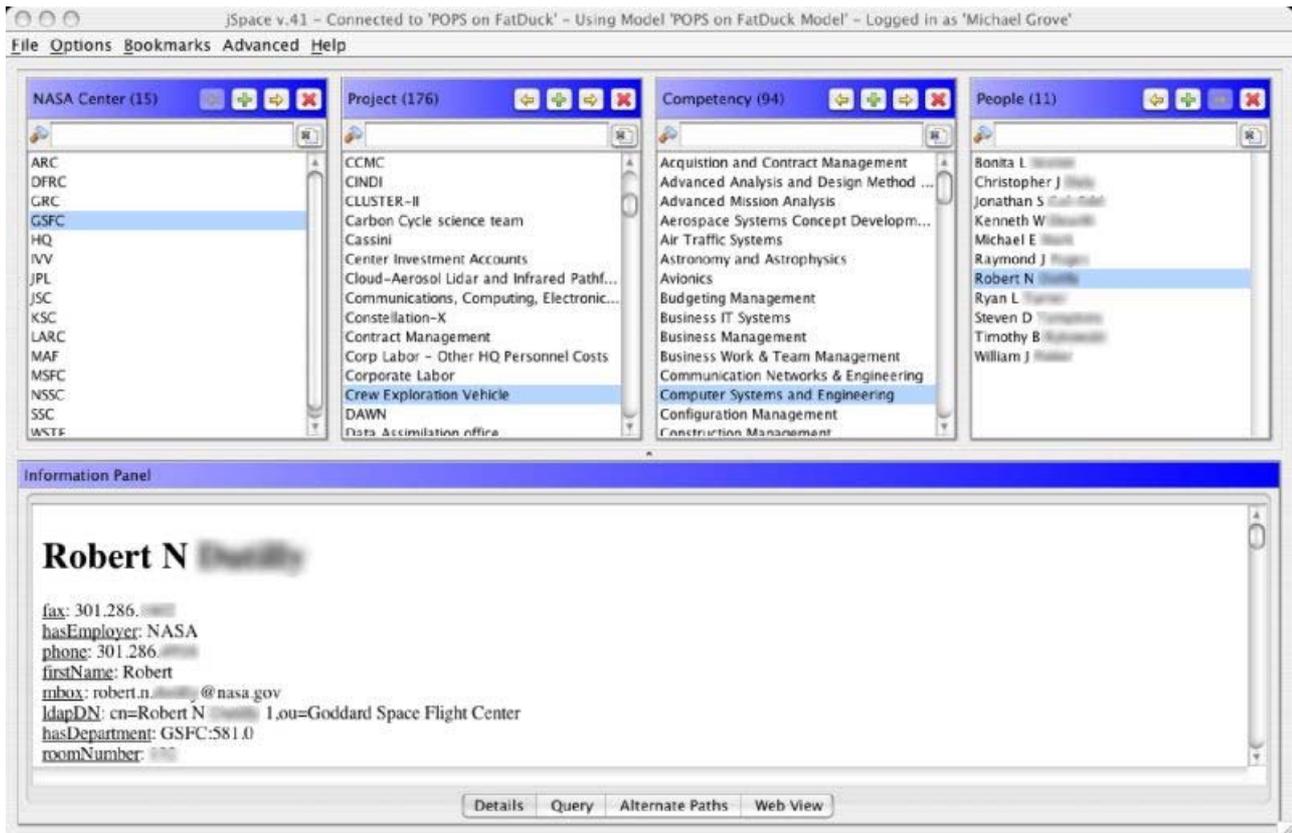


Figure 1: Browsing to find a Person in jSpace, showing the Details View (a larger version of the figure is also available)

The most interesting custom plug-in is a social network visualizer, which displays perspectival relationships between the user and other people with whom they are connected by some NASA relation (Figure 1). It displays the social network between the user and people who work in the same department, people who work on the same projects, or people with the same skill sets and competencies. POPS also includes a know-who function that, given the current user and a target person (Figure 2) who's been identified in POPS, will return an intermediate person who is related to both the user and the target person. Rather than try to change NASA's "rolodex culture", the POPS team decided to enhance and support it by helping project staffers find people with whom they could talk about candidates, their abilities, interests, qualifications, etc. POPS effectively extends the network of contacts for its users by suggesting "intermediate nodes" in the social network.

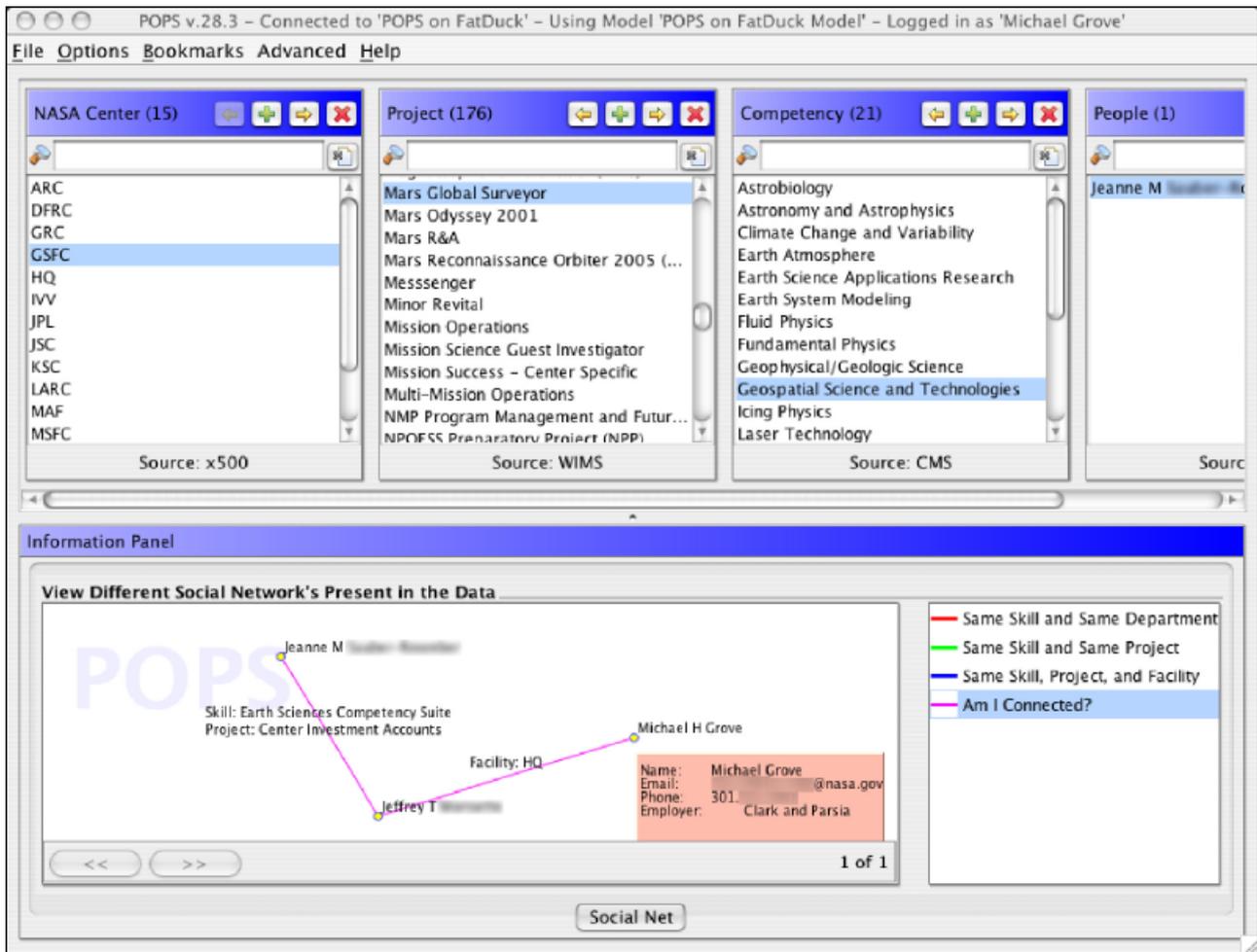


Figure 2: The "Know-Who" function of the Social Network plug-in

POPS also contains an alternate paths function for finding sets of individuals with similar characteristics. Imagine you are staffing a new NASA project and you have found all the propulsion engineers who worked on the Shuttle at Goddard. Now you would like to find other projects these engineers have also worked on. Using the alternate paths functionality, you can easily identify all of the projects these engineers have worked on. You could find just as easily all the people at Goddard who worked on the Shuttle by relaxing the current constraint that they must be propulsion engineers. An example of the Alternative Paths capability is shown in Figure 3.

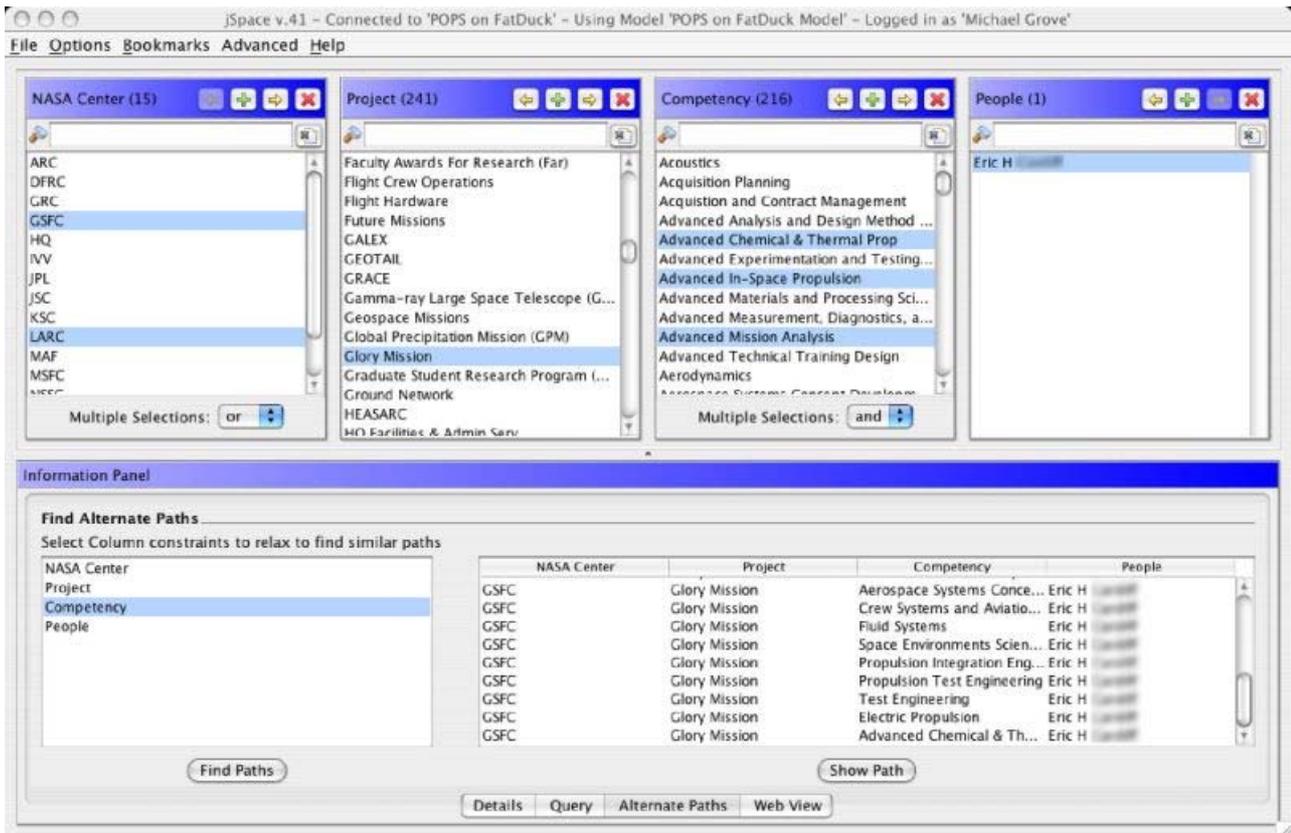


Figure 3: The Alternate Paths interface (a larger version of the figure is also available)

Using POPS it is also possible to connect to related external data on the Web. Figure 4 show how the UI can be used to connect to information about the Hubble Space Telescope in Wikipedia.

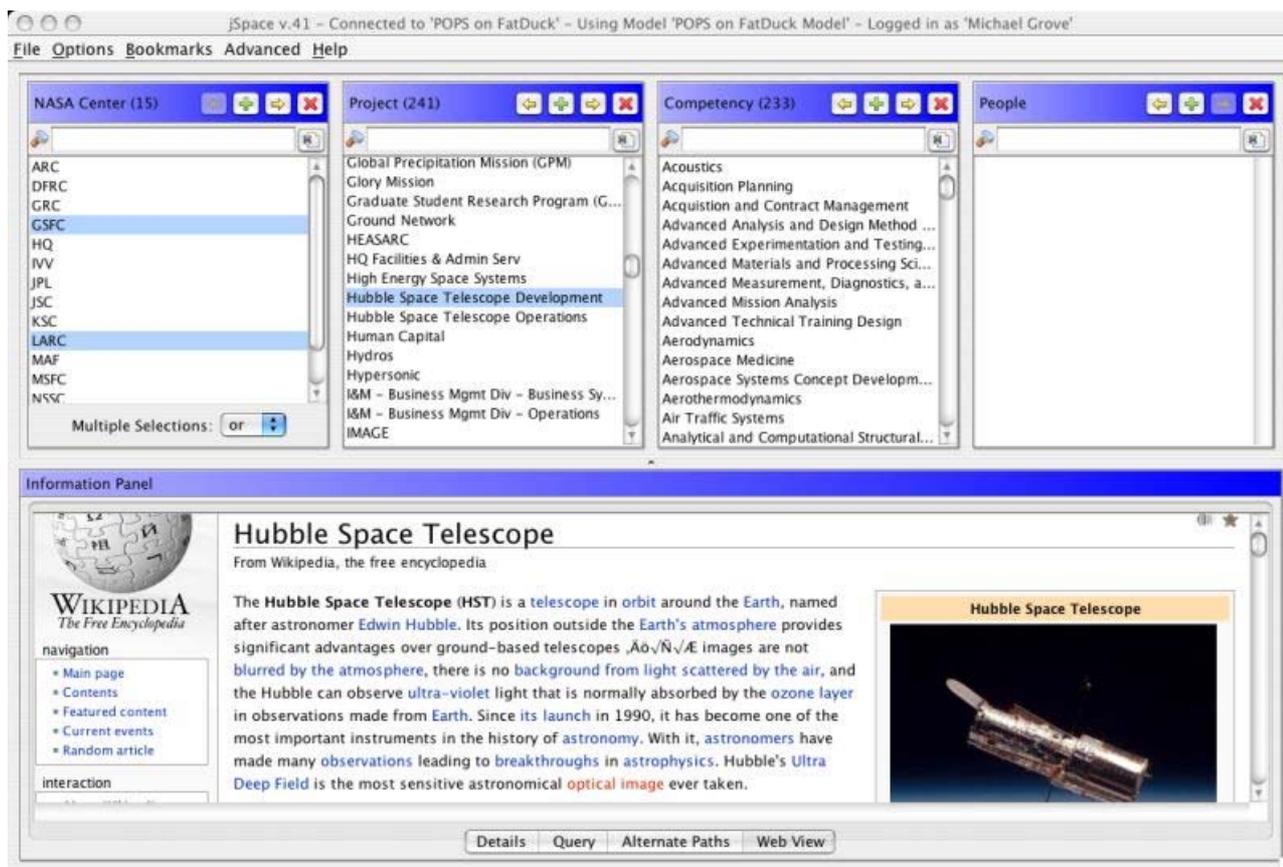


Figure 4: The Web View, which integrates existing web content into the jSpace browser (a larger version of the figure is also available)

Conclusions

By aggregating and linking information that already existed at NASA, leaving it undisturbed *in situ*, and providing an easy to use and learn interface, POPS provides not only an expertise location service, but also lightweight information integration infrastructure that will be reused to build other applications that are essentially reducible to information integration problems. Using RDF as the exchange and modeling format, together with W3C standards for RDF query and data access, NASA is taking advantage of existing investment in Web infrastructure to share and reuse information across disparate data sources and geographically distinct facilities.

By modeling not only data sources, but also their integration, in RDF, NASA gets a standard basis upon which to provide additional integrations, analysis services, and potential for a "semantic upshift" to OWL DL and related technologies.

By deploying all of these parts and pieces over lightweight RESTful web services, NASA gets rapid integration of loosely coupled and reusable infrastructure pieces.

Key Benefits of Using Semantic Web Technology

- A lightweight information integration infrastructure
- The rapid integration of loosely coupled infrastructure
- Ability to share and reuse information across disparate data sources and geographies
- A consistent data model to build upon
- The ability to reuse a powerful user interface to data