SAND2009-7698C

saple

Sandia Advanced Personnel Locator Engine

A new way to search for Sandia people and organizations

November 19, 2009

Mike Procopio, Ph.D. mjproco@sandia.gov
Cara W. Corey cwcorey@sandia.gov

Sandia Unclassified Unlimited Release SAND #: 2009-7698C

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

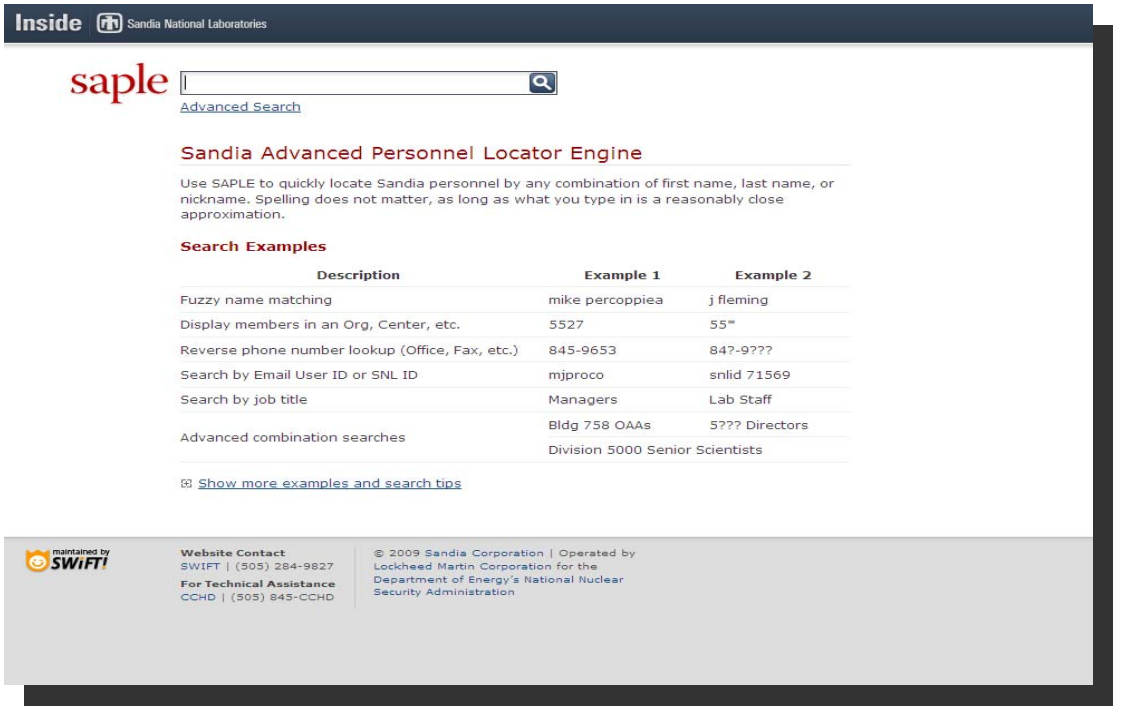
Significant Initiative to Improve Search at Sandia

- **Top-down directives, support, and funding to improve the way we organize and retrieve information at Sandia**
  - **Support at the CIO office level**
- **Modernization of search capabilities**
  - **General intranet search**
  - **Directory search**
  - **Related: Directory "Landing" page, etc.**
- **Leverage modern methods**
  - **Computer science, algorithms focus**

Sandia National Laboratories

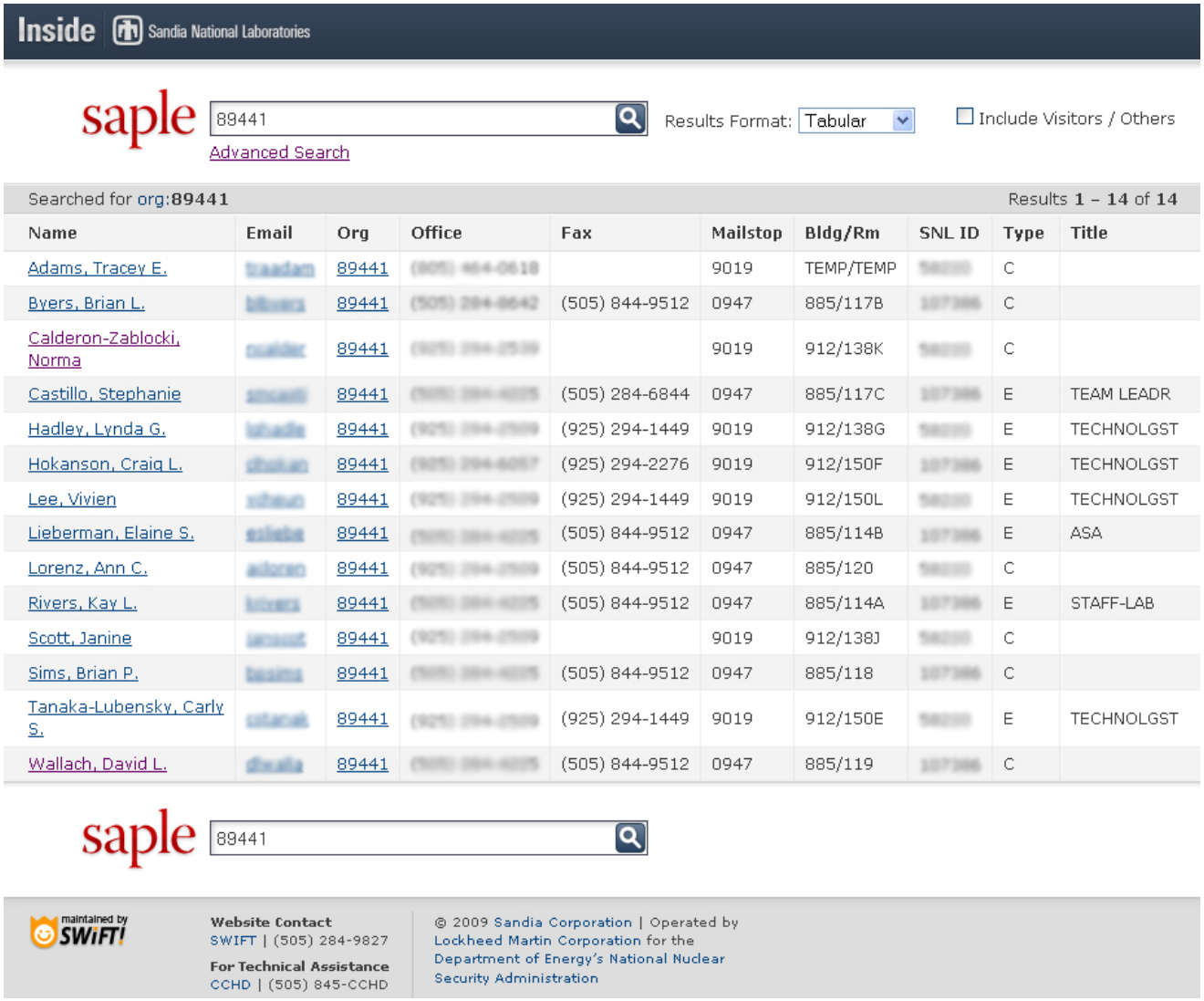
In Active Deployment

- **SAPLE represents an early success with search infrastructure modernization**
- **10,000 search queries daily**



Sandia National Laboratories

Tabular View



Inside | Sandia National Laboratories

saple 89441 Results Format: Tabular ☐ Include Visitors / Others

Advanced Search

Searched for org:89441 Results 1 – 14 of 14

Name	Email	Org	Office	Fax	Mailstop	Bldg/Rm	SNL ID	Type	Title
Adams, Tracey E.	traadam	89441	(805) 464-0618		9019	TEMP/TEMP	58233	C	
Byers, Brian L.	blbyers	89441	(505) 284-6642	(505) 844-9512	0947	885/117B	107386	C	
Calderon-Zablocki, Norma	ncalder	89441	(925) 294-2539		9019	912/138K	58233	C	
Castillo, Stephanie	smcast	89441	(505) 284-4035	(505) 284-6844	0947	885/117C	107386	E	TEAM LEADR
Hadley, Lynda G.	lghadle	89441	(925) 294-2509	(925) 294-1449	9019	912/138G	58233	E	TECHNOLGST
Hokanson, Craig L.	chokan	89441	(925) 294-6057	(925) 294-2276	9019	912/150F	107386	E	TECHNOLGST
Lee, Vivien	vlheun	89441	(925) 294-2509	(925) 294-1449	9019	912/150L	58233	E	TECHNOLGST
Lieberman, Elaine S.	eslieb	89441	(505) 284-4035	(505) 844-9512	0947	885/114B	107386	E	ASA
Lorenz, Ann C.	adoren	89441	(925) 294-2509	(505) 844-9512	0947	885/120	58233	C	
Rivers, Kay L.	krivers	89441	(505) 284-4035	(505) 844-9512	0947	885/114A	107386	E	STAFF-LAB
Scott, Janine	janscot	89441	(925) 294-2509		9019	912/138J	58233	C	
Sims, Brian P.	bpsims	89441	(505) 284-4035	(505) 844-9512	0947	885/118	107386	C	
Tanaka-Lubensky, Carly S.	cstanak	89441	(925) 294-2509	(925) 294-1449	9019	912/150E	58233	E	TECHNOLGST
Wallach, David L.	dlwalla	89441	(505) 284-4035	(505) 844-9512	0947	885/119	107386	C	

saple 89441

maintained by SWIFT

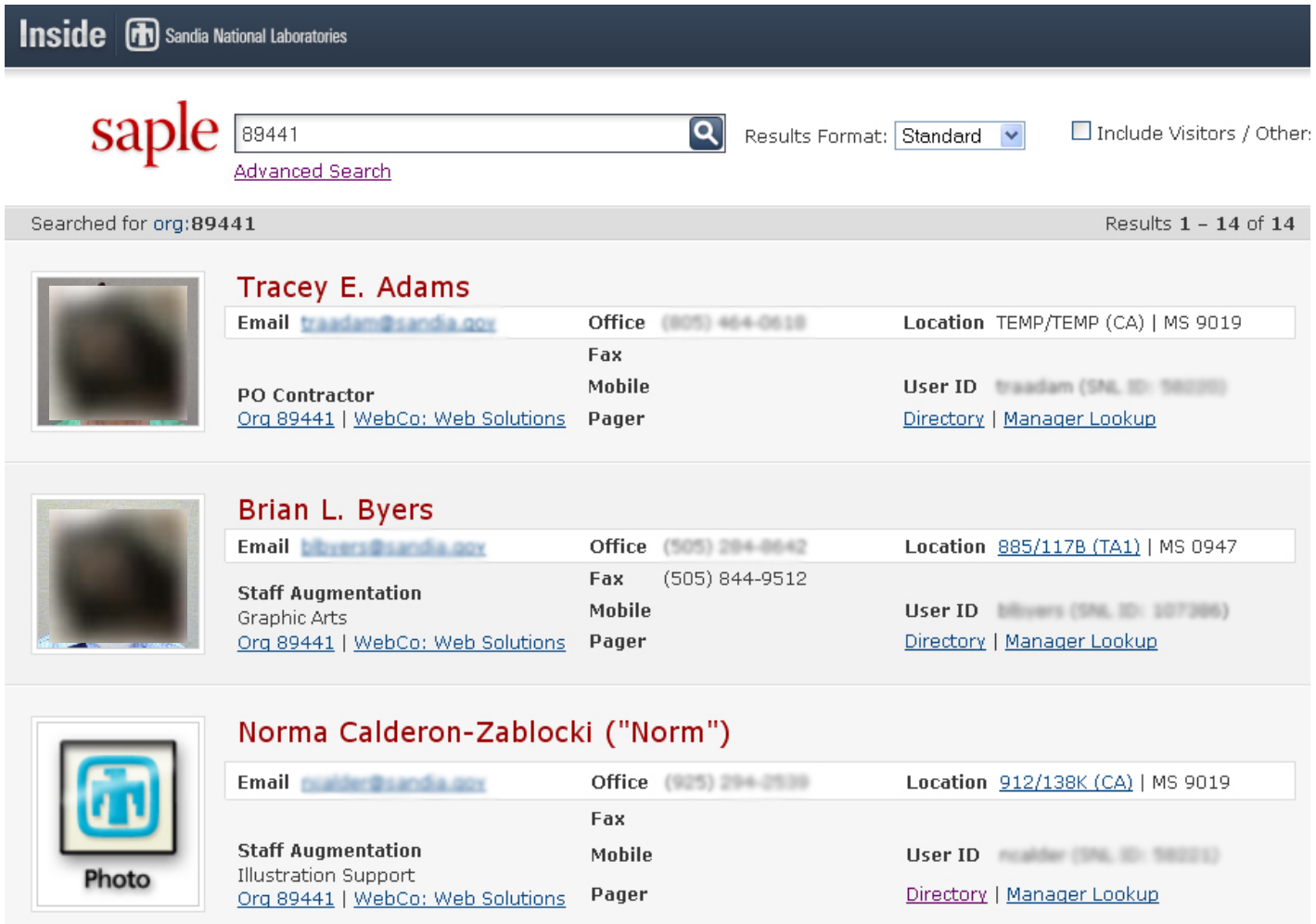
Website Contact
SWIFT | (505) 284-9827

For Technical Assistance
CCHD | (505) 845-CCHD

© 2009 Sandia Corporation | Operated by Lockheed Martin Corporation for the Department of Energy's National Nuclear Security Administration

Sandia National Laboratories

Standard View



Sandia National Laboratories

Operational Goals

- **Foundational premise: Improve the efficiency and accuracy in personnel search**
  - **Faster searches – find the right person the first time**
    - **Reduce time it takes to find someone**
  - **Recall results that were not available before**
    - **Find people that couldn't be found before**
  - **Expose advanced search functionality more easily**
    - **Natural language combination searches, like "v* building 758 oaa"**
  - **More holistic and integrative search experience**
    - **Integrate maps, user profile page, department homepage**
  - **Improve overall user experience**

Sandia National Laboratories

SAPLE Uses Modern Search Architecture

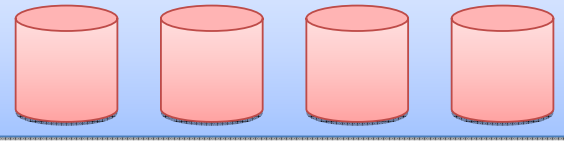
Traditional: Queries Direct to DB

SNL Directory

Search by Name
Search by Nickname
Search by Organization

```
select * from person WHERE name like 'pro%';
```

Personnel Database Tables



SAPLE: Queries Against an Index

saple

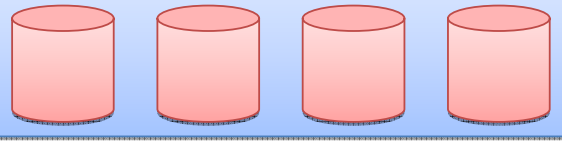
Search

Advanced Search

Artificial Intelligence and Search Algorithms

Search Index Held in Server Memory

Personnel Database Tables



Sandia National Laboratories

Three SAPLE Cornerstones

1. Inexact (approximate) string matching
2. Intelligent, compound query interpretation
3. Support of search analytics to improve results over time

Examples of inexact string matching are becoming increasingly common

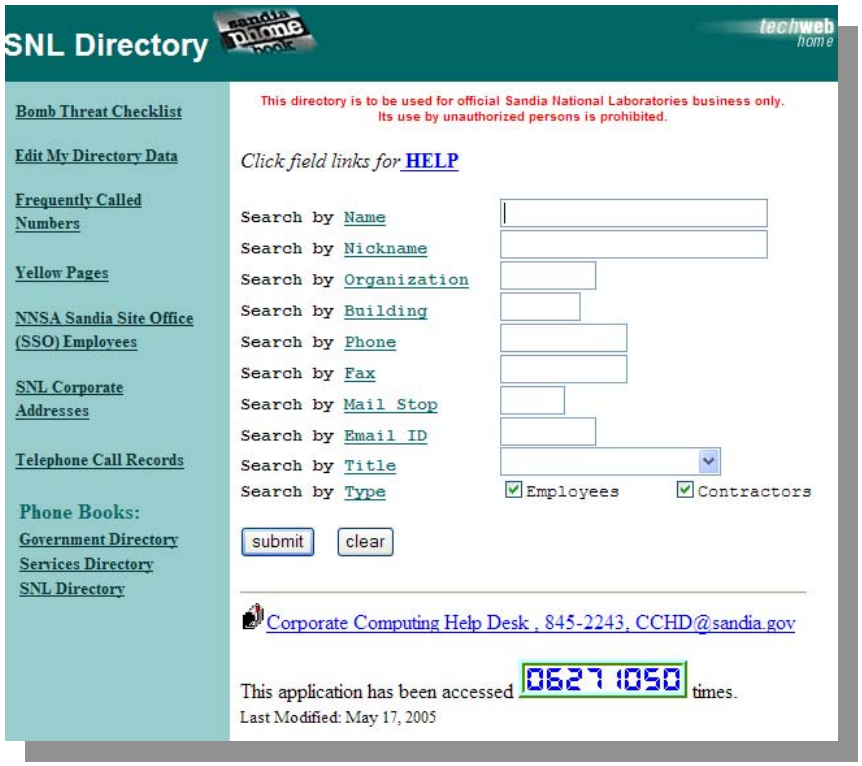
Google: “Bernalilo country metr cort”

Facebook: “Sharron Procopia”

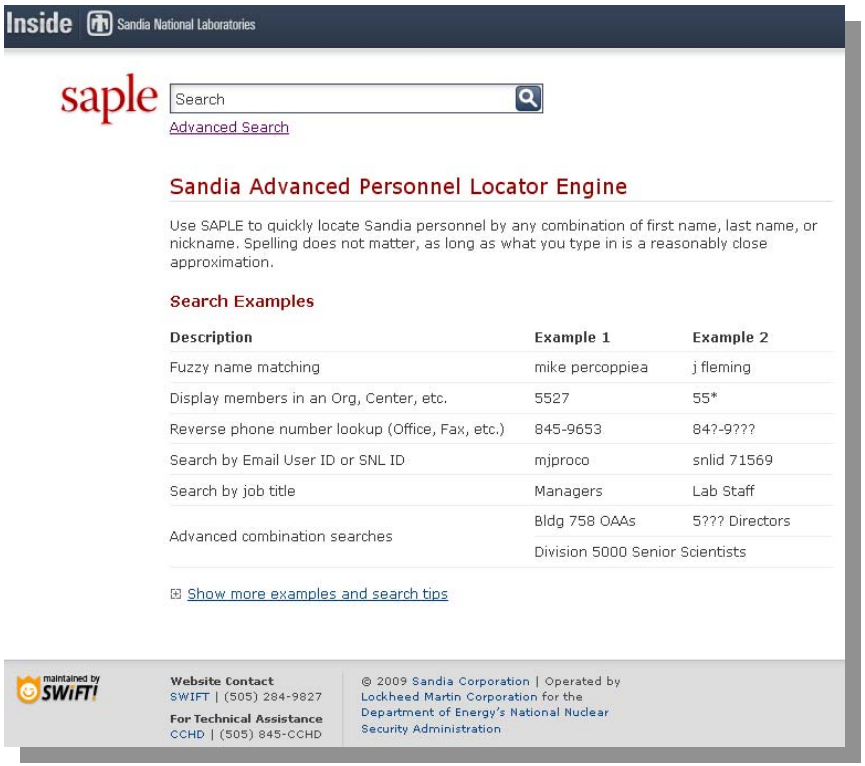
Sandia National Laboratories

Better User Experience with One Query

Traditional: User must choose where to put query



SAPLE: Query is automatically interpreted



Sandia National Laboratories

Search Analytics

- **Premise of Search Analytics**
  - **Improve search quality over time by mining rich archive data logs of query history**
- **Analyzing log data shows us:**
  - **Usage trends**
    - **Number of queries daily, peak search load times, etc.**
  - **Where people are searching**
    - **Techweb home page, re-searching in SAPLE header/footer, external application hooks, SAPLE home page, etc.**
  - **What people are searching for**
    - **Last name only queries, wildcarded queries, org roster listings, compound queries, user IDs, manager lookups, etc.**
    - **Can take name-based queries and identify percentage that would have previously failed due to misspellings**
  - **How well SAPLE is performing computationally**
    - **Average search response time**
  - **How well SAPLE is performing algorithmically**
    - **Percentage of the time people found who they looked for on the first query**

Challenges with Deploying a New Way to Search

- People are resistant to change…
  - How do we improve search and other IT infrastructure offerings without disrupting flow of business activities?
- CIO supports this:
  - We must continue to evolve our offerings and progress our capabilities … but change is uncomfortable. We must be cautious not to let vocal minority impede progress.
  - Tendency is do not improve capabilities for fear of vocal minority
- Quantifiable search improvement can help make the case

Sandia National Laboratories

Making the Case for Change

- **Existing code base and capabilities will not be sustainable forever**
  - **Legacy code, languages, paradigms, interfaces, dependencies…**
- **Assumptions and dependencies of code will change**
  - **Changes in underlying personnel database structure; e.g., imminent retirement of legacy SNL Directory application**

Lessons Learned

- (If at all possible) don't rush to deployment – ensure that the tool is thoroughly tested first
  - While a supportive user base can tolerate advances in functionality with minor glitches in the early stages, don't assume your users have endless sources of patience and good will
- Be prepared for things to appear to be wrong – SAPLE exposed underlying issues with the data (strange nicknames, compound first names, data formatting)
- Edge cases can and will come up
  - Searching for "Joe Nevada" shouldn't return "Nevada Test Site"
  - Fast, agile software development with frequent minor updates
  - Quick turnaround time to fix bugs
  - Become comfortable with incremental, minor redeployments

More Lessons Learned

- It's not easy for users to change their old habits
  - Most searches are still in the "lastname, firstname" format
- Exploring the limits of the corporate computing infrastructure
  - SAPLE is a relatively high-volume application: 10,000 queries daily
  - Each query associated with server-side algorithmic computation (e.g., string matching algorithms), which can be burdensome for a server
  - Applications have many points of failure
    - Database, Kerberos authentication, network hardware,
    - And the WebLogic production server itself with 70+ other shared applications

Still More Lessons Learned

- **Increased demands on developer skill set, and it's not a simple process to transfer application support from one developer set to another:**
  - **Simple web-app front end that crafts SQL query from discrete textbox fields, executes query, returns results in tabular form**
  - **C, Perl**

    vs.
  - **String matching algorithms, regular expression matching for query interpretation, advanced data structures, multi-threading, algorithm optimization, logging and analytics, advanced user interface with rich client interactions**
  - **Java, Java Enterprise, Servlets, XML, algorithms, database design, stored procedures, optimization**
  - **Some advanced computer science concepts**