Diving into Data
Examples from the Exploratorium

Sherry Hsi

shsi@exploratorium.edu

November 18, 2009

NSDL Annual Meeting, Washington, DC

Learning Resources Collection

Advanced Search

Search results

Driving Questions

- **Who is using our site ?**
  - What is the geographic distribution?
  - What is the education level of the audience?

- **Why are they visiting?**
  - Can we identify different behaviors for different users?

- **Can we improve our collection's discoverability?**
  - What are the search keywords that lead people to the site?
  - What content do users visit? Which resources are the most popular?
  - What is the usage and growth of ELRC?

- **What's the impact of ELRC?**
  - Are they learning STEM content and ideas?

# Data Sources

Web usage:
- **Web server log – since 2005**
- **Google Analytics (GA) – since Nov 2007**

Other:
- **Zipcodes (2 years, 2005-2007)**
- **Online survey + zipcode (4 months, Fall 2009)**
- **Search term capture**
- **Navigation Paths**
- Interviews (not yet)

Triangulation & Data Combining

Logfiles – Navigation Paths,

Google Analytics
- GeoIP

Unique Jsession ID

ZCTA
zipcode
(census)
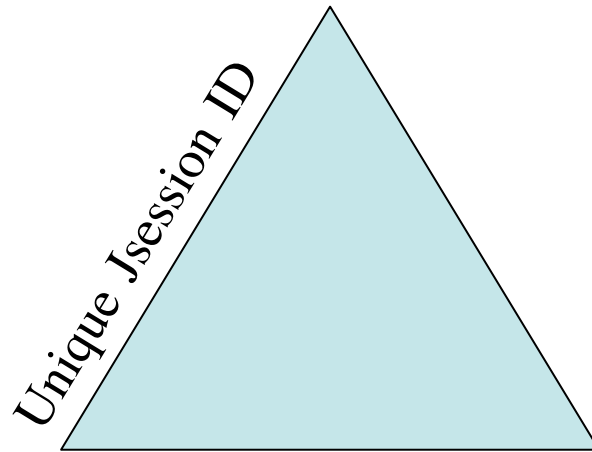
Zipcodes

**User online
survey data**

Phone #, email

**Phone
interview
follow-up**
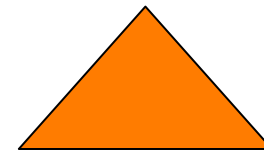
Google Analytics - visits

Country/Territory Detail:
United States

Nov 1, 2007 - Nov 4, 2009

Visits

Zoom Out
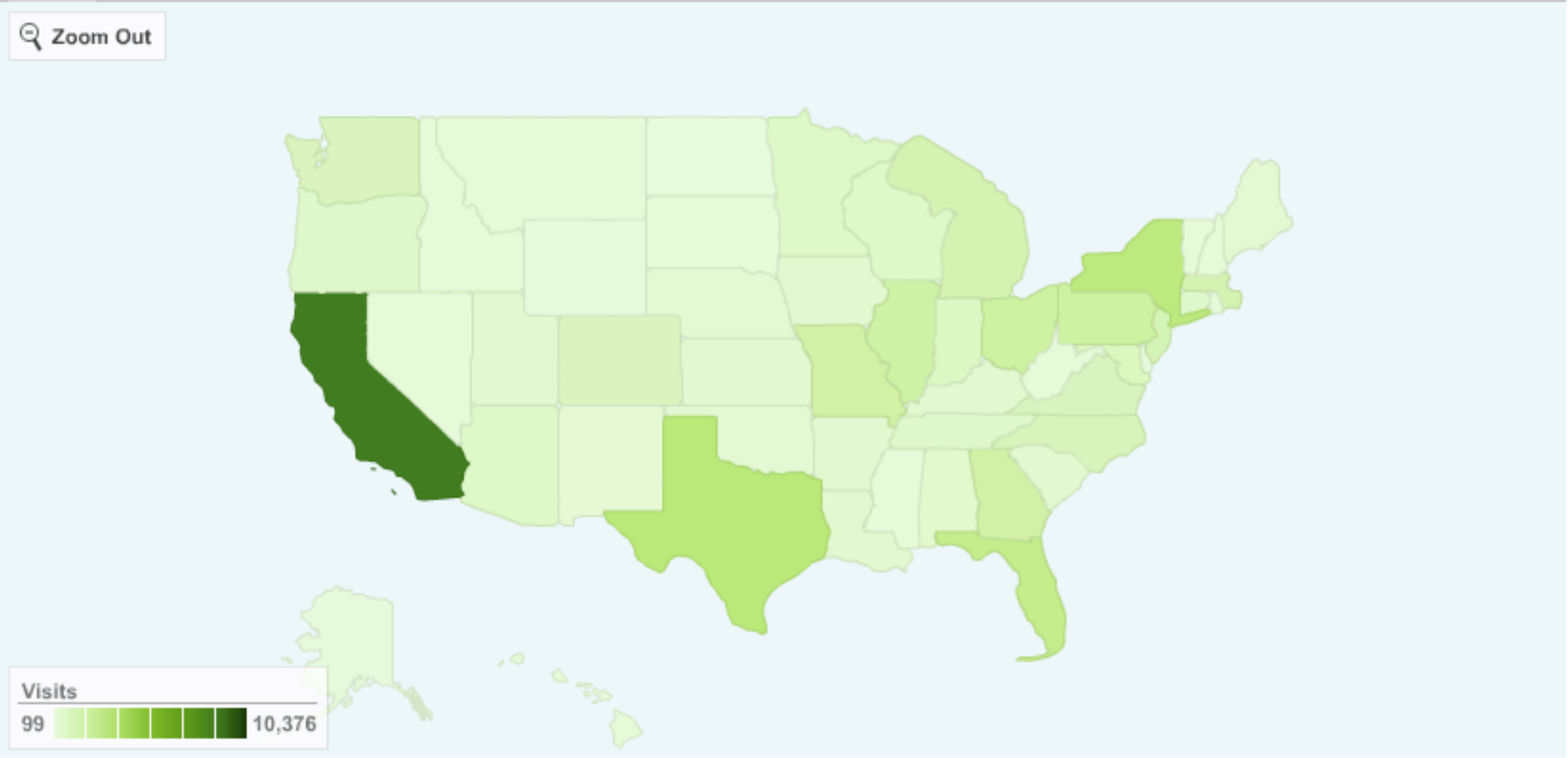
Visits
99 ▬▬▬▬▬▬▬ 10,376

This country/territory sent 56,915 visits via 52 regions

Census Data with GeoIP - # of schools and online visits
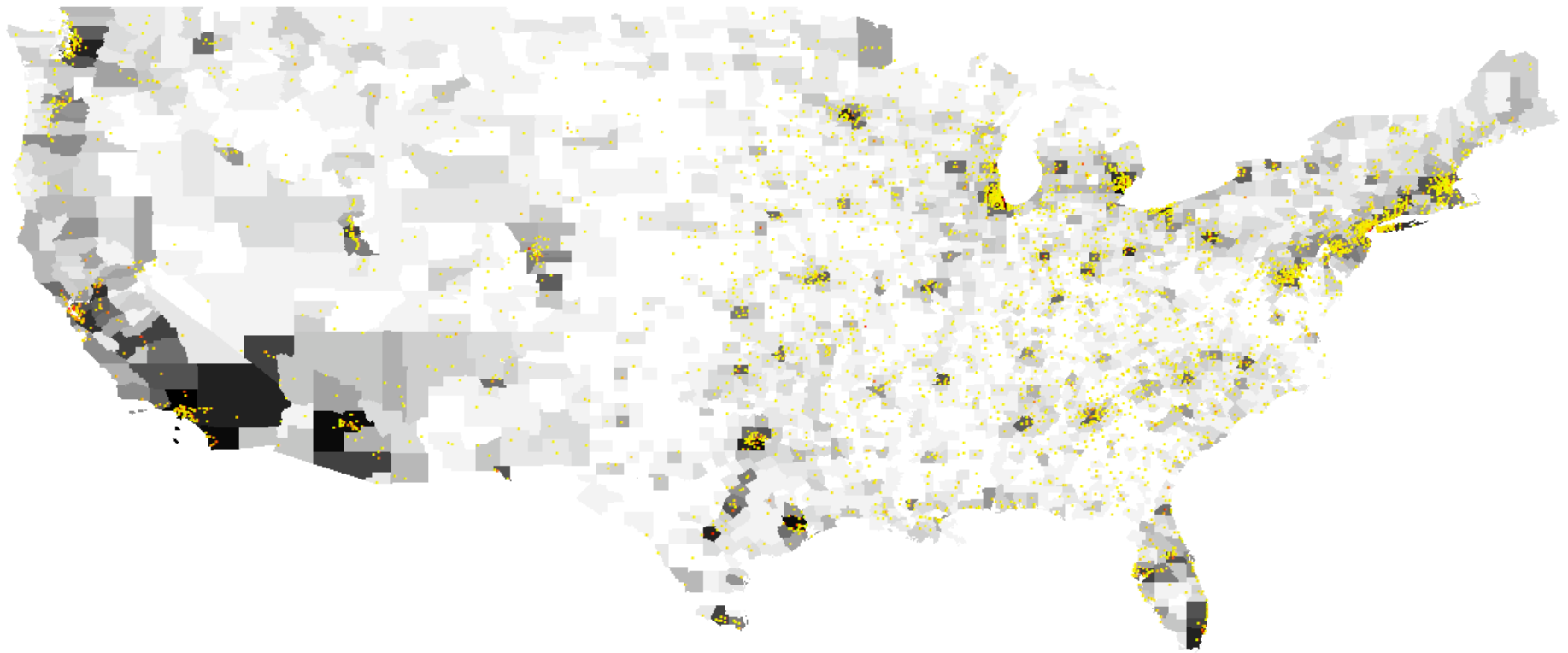
Zipcodes from survey (4 months, Fall 2009)

Unknown 7%

OUTSIDE U.S. 26%

U.S. 67%
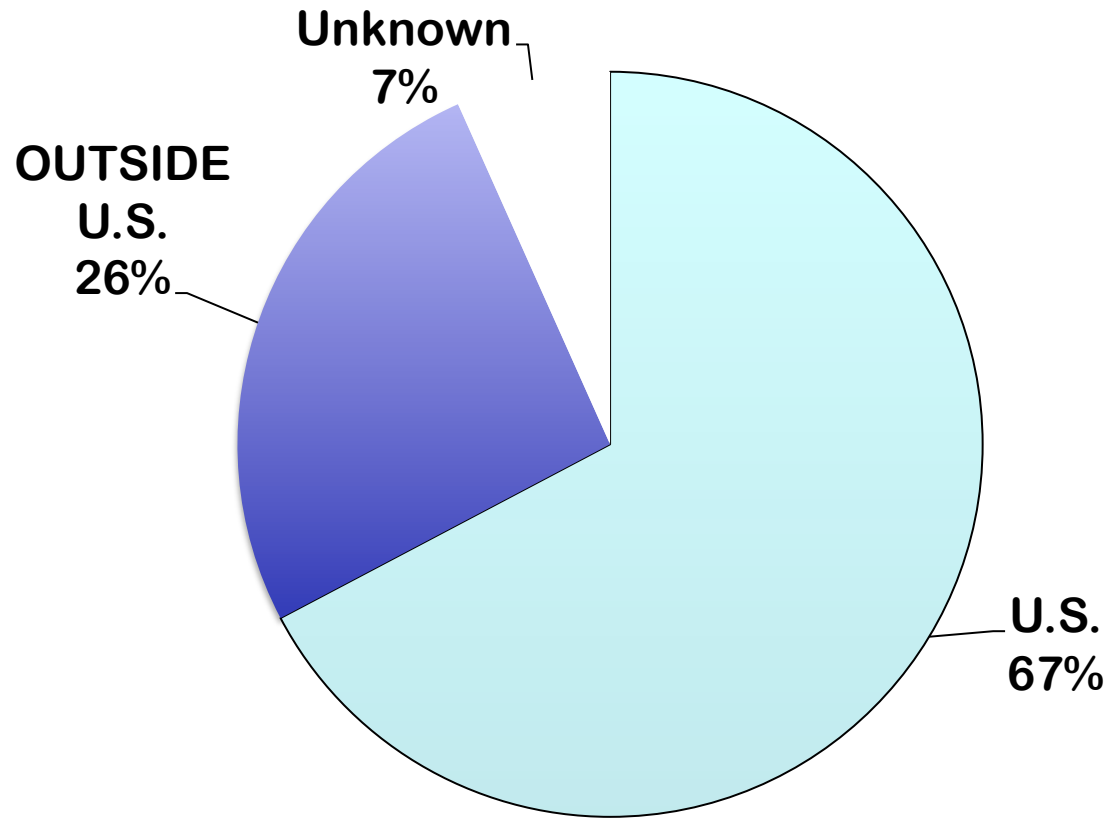
Number of Visitors by Region (Zipcodes from survey, Fall 2009)

http://www.usnaviguide.com/zipunion.htm
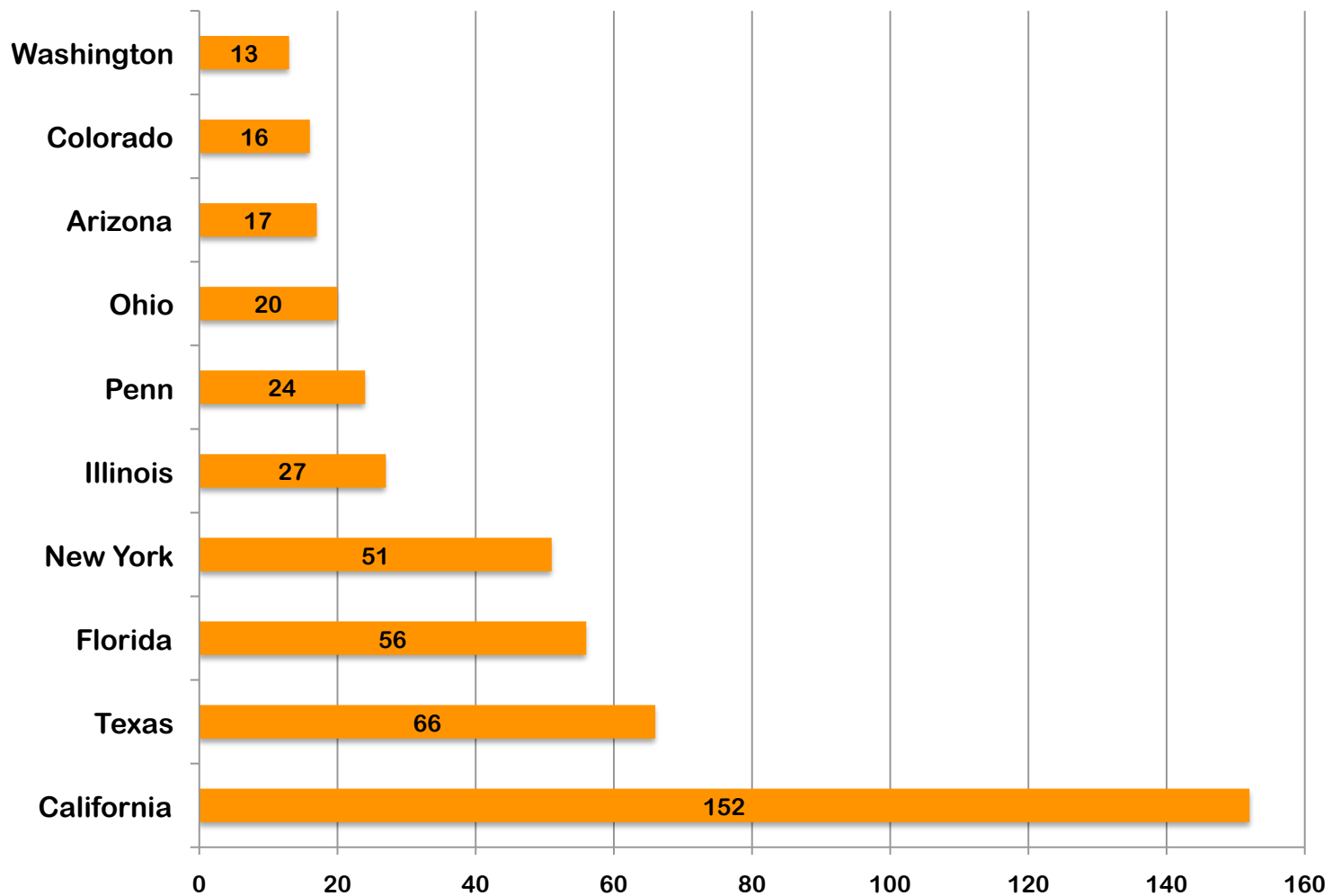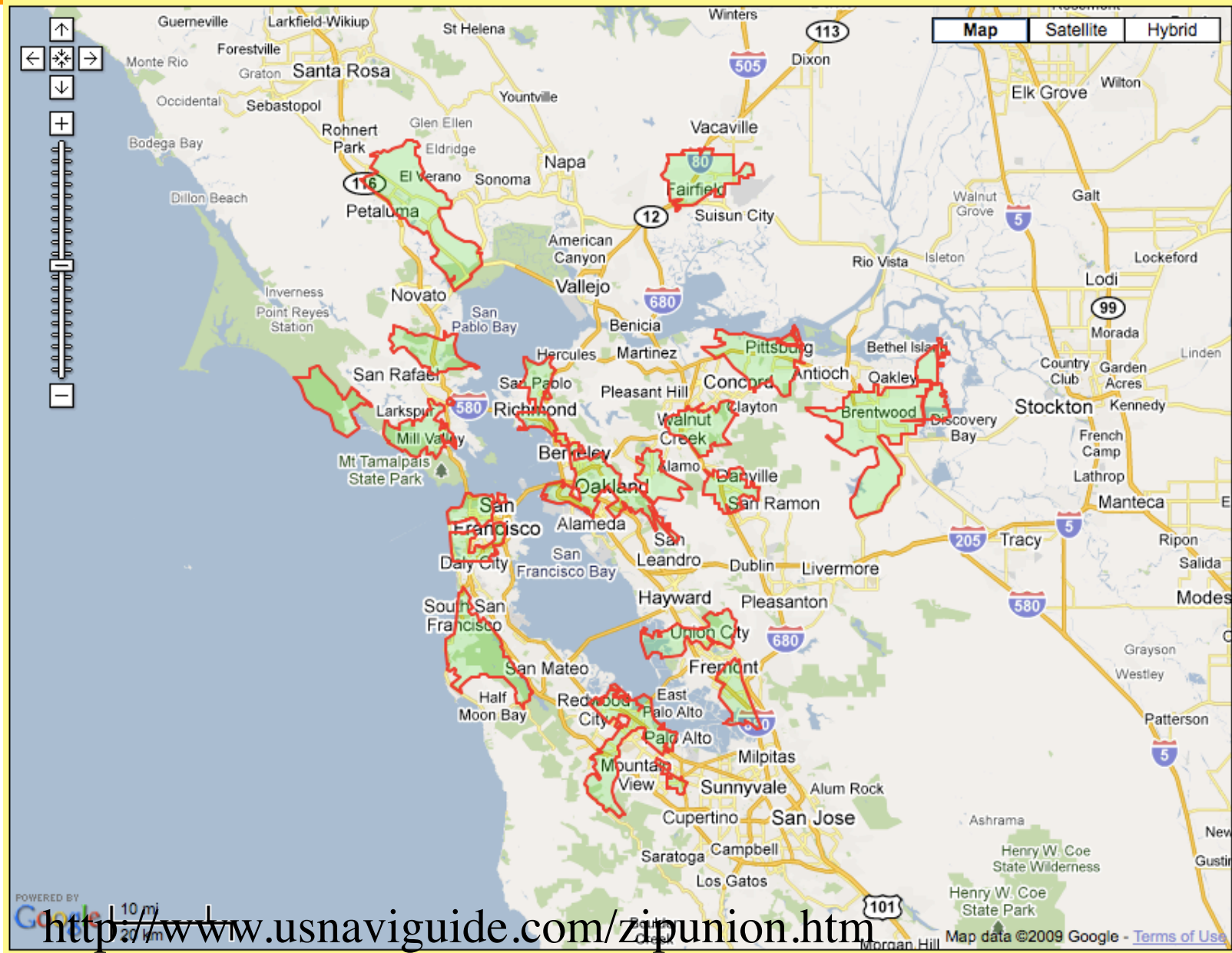
Google Analytics – Who are you?

Visitors Overview

Aug 1, 2008 - Aug 31, 2009

Visitors

| 300 | | | | | | | | | 300 |

| 150 | | | | | | | | | 150 |

Graph by:

August 4, 2008 | September 9, 2 | October 15, 20 | November 20, : | December 26, : | January 31, 20 | March 8, 2009 | April 13, 2009 | May 19, 2009 | June 24, 2009 | July 30, 2009
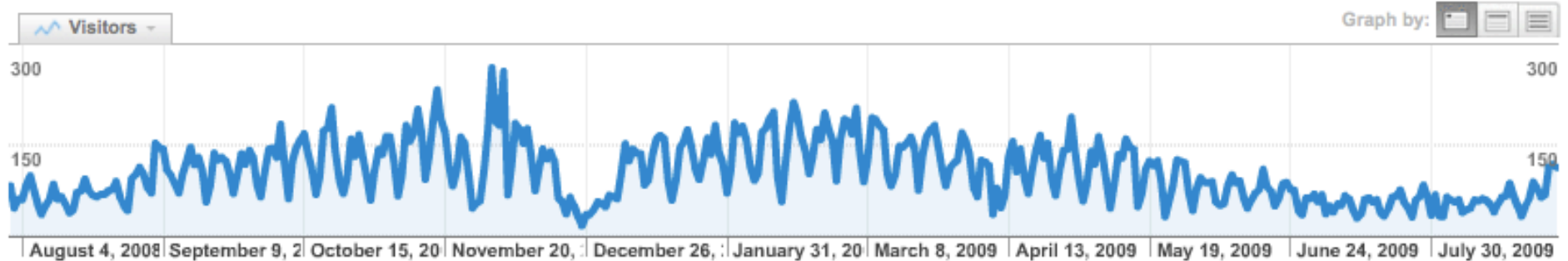
38,922 people visited this site

44,799 Visits

38,922 Absolute Unique Visitors

143,422 Pageviews

3.20 Average Pageviews

00:02:20 Time on Site

57.99% Bounce Rate

86.37% New Visits

Visitor Segmentation

Visitors Profile: languages, network locations, user defined

Browser Profile: browsers, operating systems, browser and operating systems, screen colors, screen resolutions, java support, Flash

Map Overlay
Geolocation visualization

Try Google Ad Planner.
Understand the demographics and behaviors of your website's visitors.

LogFile (Tomcat server)

202.164.195.28 - - [06/Sep/2009:01:26:23 -0800] "GET /nsdl/images/explo_corner.gif HTTP/1.0" 200 210 "http://nsdl.exploratorium.edu/nsdl/welcome.do" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.13) Gecko/2009073022 Firefox/3.0.13" 7974C9D96F3F640A951CBFB9BB4FF78B

202.164.195.28 - - [06/Sep/2009:01:26:23 -0800] "GET /nsdl/images/nsdl_35ltblu.gif HTTP/1.0" 200 1863 "http://nsdl.exploratorium.edu/nsdl/welcome.do" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.13) Gecko/2009073022 Firefox/3.0.13" 7974C9D96F3F640A951CBFB9BB4FF78B

msnbot-65-55-207-74.search.msn.com - - [06/Sep/2009:01:26:40 -0800] "GET /nsdl/search_results.vm?qid=SAQ-11&itemsPerPage=10&index=160 HTTP/1.0" 200 8035 "-" "msnbot/2.0b (+http://search.msn.com/msnbot.htm)" 9808D93AA28EE37BB43D6EBA65643F5E

202.164.195.28 - - [06/Sep/2009:01:26:52 -0800] "GET /nsdl/search.do?category=Curricular+Area-Earth+and+Space+Science-Astronomy HTTP/1.0" 200 38130 "http://nsdl.exploratorium.edu/nsdl/welcome.do" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.13) Gecko/2009073022 Firefox/3.0.13" 7974C9D96F3F640A951CBFB9BB4FF78B

202.164.195.28 - - [06/Sep/2009:01:26:52 -0800] "GET /nsdl/css/exploratorium.css HTTP/1.0" 200 8206 "http://nsdl.exploratorium.edu/nsdl/search.do;jsessionid=7974C9D96F3F640A951CBFB9BB4FF78B?category=Curricular+Area-Earth+and+Space+Science-Astronomy" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.13) Gecko/2009073022 Firefox/3.0.13" 7974C9D96F3F640A951CBFB9BB4FF78B

202.164.195.28 - - [06/Sep/2009:01:26:53 -0800] "GET /nsdl/getThumbnail.do?id=12085 HTTP/1.0" 302 - "http://nsdl.exploratorium.edu/nsdl/search.do;jsessionid=7974C9D96F3F640A951CBFB9BB4FF78B?category=Curricular+Area-Earth+and+Space+Science-Astronomy" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.13) Gecko/2009073022 Firefox/3.0.13" 7974C9D96F3F640A951CBFB9BB4FF78B

202.164.195.28 - - [06/Sep/2009:01:26:53 -0800] "GET /nsdl/getThumbnail.do?id=12104 HTTP/1.0" 302 - "http://nsdl.exploratorium.edu/nsdl/search.do;jsessionid=7974C9D96F3F640A951CBFB9BB4FF78B?category=Curricular+Area-Earth+and+Space+Science-Astronomy" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.13) Gecko/2009073022 Firefox/3.0.13" 7974C9D96F3F640A951CBFB9BB4FF78B
202.164.195.28 - - [06/Sep/2009:01:26:53 -0800] "GET /nsdl/getThumbnail.do?id=12093 HTTP/1.0" 302 - "http://nsdl.exploratorium.edu/nsdl/search.do;jsessionid=7974C9D96F3F640A951CBFB9BB4FF78B?category=Curricular+Area-Earth+and+Space+Science-Astronomy" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.13) Gecko/2009073022 Firefox/3.0.13" 7974C9D96F3F640A951CBFB9BB4FF78B
202.164.195.28 - - [06/Sep/2009:01:26:53 -0800] "GET /nsdl/getThumbnail.do?id=12075 HTTP/1.0" 302 - "http://nsdl.exploratorium.edu/nsdl/search.do;jsessionid=7974C9D96F3F640A951CBFB9BB4FF78B?category=Curricular+Area-Earth+and+Space+Science-Astronomy" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.13) Gecko/2009073022 Firefox/3.0.13" 7974C9D96F3F640A951CBFB9BB4FF78B

Tracking a search path

8E333935DB2D7C6CE71C6AD45C27950C	anonymous	science	Sun, September 6, 2009 3:16:26 pm PDT
6E9BF4036C3B94BAB9AA6F051909F966		energy	Sun, September 6, 2009 3:38:43 pm PDT
D9661FA083299948FA0B374371BA008F		time warp	Sun, September 6, 2009 6:44:41 pm PDT
E285B33D674ADDEA5303D2F983BC38DD		rocks and minerals	Sun, September 6, 2009 8:15:52 pm PDT
88DA30D12C150C3BAAA660291F5C6380		silica dioxide	Sun, September 6, 2009 8:29:34 pm PDT
4E8F2C16863A631AB562B26ACFDBBA04		resonant rings	Mon, September 7, 2009 12:40:21 am PDT
F585E40DDD101A56059997F606ED3998		how a camera is made	Mon, September 7, 2009 9:25:21 am PDT
398C7350EBDD8FF0058FA6D7B192DB76		geothermal energy	Mon, September 7, 2009 9:26:00 am PDT
7B787F56474AE0C8B8A71E80434A0698		pendulum	Mon, September 7, 2009 10:13:33 am PDT
D0040080E2C14008B11194BFBF9156F1	anonymous	physics of bikes	Mon, September 7, 2009 12:24:53 pm PDT
340E6006A7801490DDAFB9D1156A5F1F	anonymous	grade 4	Mon, September 7, 2009 1:35:17 pm PDT
DCBDFDB1AE98C64A44F3E1A9EAA6B4DE		osmosis and diffusion	Mon, September 7, 2009 3:38:45 pm PDT
55D4794DA823B7D896209195712D737D		space	Mon, September 7, 2009 4:31:21 pm PDT
8FADE48A2F8FF241132B4DE76803F1B9		science	Mon, September 7, 2009 5:26:40 pm PDT
1560EEE447D7773667398E952EB89B58		tonic water	Mon, September 7, 2009 7:55:01 pm PDT
E298FD25E67F62CF7A9282A8341C7684		matter	Tue, September 8, 2009 8:18:40 am PDT
E7DC47345377904A5C6D92849F235EC2		kindergarten	Tue, September 8, 2009 10:08:27 am PDT

Search Activity for JSESSIONID: 79A712A0C87A5E2C16BEB8D9C2B12700

| Term | URL Visted | # Results Returned | Timestamp | Topics chosen | Advanced Search Options | | Results per page chosen |
					Grades chosen	Resource Type chosen	
earth	Initialize search from: http://nsdl.exploratorium.edu /nsdl/advanced_search.vm?qid=	81	Fri, August 28, 2009 10:10:05 am PDT	Curricular Area-Earth and Space Science	6-8 9-12 Postsecondary	Activity Article/Text Image Professional Development Materials	5
	http://nsdl.exploratorium.edu /nsdl/showRecord.do?id=10986		Fri, August 28, 2009 10:10:30 am PDT				
	id=10917		Fri, August 28, 2009 10:10:55 am PDT				
	http://nsdl.exploratorium.edu /nsdl/showRecord.do?id=12110		Fri, August 28, 2009 10:11:40 am PDT				
	id=12110		Fri, August 28, 2009 10:12:07 am PDT				
acid rain	Initialize search from: http://nsdl.exploratorium.edu /nsdl/showRecord.do?id=12110	6	Fri, August 28, 2009 10:14:06 am PDT				5
	http://nsdl.exploratorium.edu /nsdl/showRecord.do?id=10938		Fri, August 28, 2009 10:14:24 am PDT				
	http://nsdl.exploratorium.edu /nsdl/showRecord.do?id=10985		Fri, August 28, 2009 10:14:53 am PDT				
	id=10985		Fri, August 28, 2009 10:15:42 am PDT				
	http://nsdl.exploratorium.edu						

Online survey – Who are you?

Submitter Roles	Count	%
K-12 Teacher	492	49
Other	127	12.6
Undergraduate Student	86	8.6
Instructor or Professor	57	5.7
Graduate Student/Post-doc	48	4.8
Afterschool Educator	41	4.1
Homeschooling Educator	38	3.8
Museum Educator	32	3.2
K-12 Librarian	20	2
K-12 Professional Developer	20	2
R&D Professional	18	1.8
Community-based program educator	17	1.7
K-12 Administrator	9	0.9

54% K12

19% Higher Ed

Why are you visiting?

Visitor Purpose	Count	%
Lesson planning.	490	48.8
Searching for a particular activity or resource.	257	25.6
No reason. Just browsing and being curious.	145	14.4
Other	72	7.2
Showing this to someone else.	29	2.9

OTHER PURPOSE: (write in)

"I have a test on sat. day and I am trying to study. For the GED test."

"A parent happy to find project ideas to do in the coming winter days at home."

"investigating color effects for dichroic glass"

"11- 18 science high school teacher in UK. No idea what that is in grades."

"Currently an education-major, just doing some research"

"i have 12 grandchildren who i visit often. i also volunteer with young children."

Google Analytics - Search Terms

Top keywords: August 1 – November 1, 2009

1. elodea leaf cells
2. elodea leaf cell
3. continental drift animation
4. "roto copter"
5. doppler effect sound file
6. how to make a jitterbug
7. salt sculpture
8. solar system
9. volvox globator

ELRC entries - Search Terms

- **TOP TEN SCIENCE**

students 5887

energy 5629

explain 5598

scientific 5105

identify 4695

earth 4521

science 4109

data 3349

organisms 3347

water 2778

- **TOP TEN MATH**

numbers 6865

solve 6545

problems 6310

data 5324

number 5153

students 5130

identify 4009

mathematical 3557

determine 3411

apply 3271

What web traffic won't tell us

- Why don't they come?
- How effective were the digital resources in meeting the needs of the end user?
- Did the user learn any science?

What EDM might tell us…

- What types of users (educators) are there?
- What are they doing on our site?
- How are they using digital resources?
- How should a digital library be changed to improve user experience?

Next steps 2009-2010

- Applying EDM cluster analyses from IA to ELRC
- Doing the triangulation and data verification
- Conducting the interviews from online users
- More analyses of search behaviors and path analyses

Lessons Learned

• Data cleaning is laborious. Set up data cleaning head of time if possible before dumping into log files.

• Iterative refinement – "keep tasting the batter" (or make sure you are collecting what you think you are collecting)

• Looking at online behaviors to infer learning

• We are doing research, not evaluation…however, try to identify effects and define impact of learning – will be different for different stakeholders.

Acknowledgments

Educational Data Mining

http://edm.usu.edu
http://www.exploratorium.edu/