

What are the procedures for assessing the outcomes of interactive informal learning Web sites?

How does the process of evaluating museum exhibits compare with the process of evaluating informal learning Web sites?

How much do we really know about our audience and what they are doing at our Web sites?

**WEB DESIGNS FOR
INTERACTIVE
LEARNING
CONFERENCE**

June 15-18, 2005
Ithaca, New York



Web Site Evaluation

Assessing the Outcomes of Interactive Web Sites 2

Saul Rockman, President, ROCKMAN *ET AL*

Evaluating Museum Exhibits and Online Programs 12

Minda Borun, Museum Solutions; Director of Research and Evaluation, The Franklin Institute Science Museum

Who's Out There and What Are They Doing Anyway? 16

A Personal Journey Through Metricland

Rob Semper, Executive Associate Director, Exploratorium



*I don't necessarily agree
with everything I say.*

Marshall McLuhan

Assessing the Outcomes of Interactive Web Sites

Saul Rockman, President, ROCKMAN *ET AL*

I am going to inundate you with a lot of ideas and information. Some of this you will already know and may already be doing. Some of this you may never have heard of, or you may have heard of it but never tried it. I hope this generates discussion, and I hope it generates ideas that you can take back with you and use.

Formative Assessment

I'm only going to spend a little time on formative assessment because most of you already do it. I just want to make sure that some of these ideas are out there so that if you haven't tried them, maybe you will try them in the future.

Formative Assessment

- Concept Testing (Focus Groups, Surveys, etc.)
- Paper Mock-Up Usability Tests
- Beta Tests
- Usability / Navigation (think-alouds, tasks, universal design)
- Appeal / Attitude
- Comprehension

I think the notion of concept testing is critical. This is the front-end piece that asks, are we doing the right kinds of things for the kind of people who are going

to come to our site?

The process of doing a paper mock-up is a tough one for a lot of designers. Actually, sending anything out but the finished product for evaluation is a problem for a lot of designers. They'll say, "It's not ready yet! The color isn't right!" I don't care, and the people who are going to test it this early in the game don't care. From my point of view as an evaluator, if it's clearly not finished, I can get people to tell me a lot about how they think it ought to be finished; if it's perfect, it's too late.

I think most of you already deal with usability and navigation issues in formative assessment, and there are a variety of ways of doing that. I want to reiterate what others at this conference have already said about the notion of ensuring you keep universal design features in mind as you create your site. That becomes increasingly important as more and more people with various ways of acting and various ways of thinking get involved on the Web.

One of the things that a lot of people don't spend much time doing is studying comprehension. Can people understand what is on your site? Can they read and understand the words? Do they have the vocabulary? As we look at sites that have something to do with learning and museum visitors, we can see that they already incorporate some vocabulary words that people don't get. As you begin your formative testing, make sure you do something to get those involved with Web site development to think a little more about what language they are using.

Are your Web visitors understanding the relationship among objects on a screen? Could they narrate the

screen for you? Can they tell you what's up there on the screen, and can they tell you what you're supposed to do with it? That's different from just asking visitors, "Tell me what's wrong with the Web site."

Where's the Proof?

One of the things that often frustrates me is when people come up with the wrong metaphor. A lot of that may stem from the fact that many people aren't reading the great literature of the world or following the history of the world. People have narrow views of things. I often find myself listening to someone who is misstating an aphorism, and this is one that drives me nuts:

*The proof is in the pudding.**

(*not Bohr's plum pudding model)

The proof is *not* in the pudding. If it were, you'd be swimming through the pudding trying to get to the truth. That is the way a lot of people approach their Web sites and projects—they look at the pudding.

The real aphorism is as follows:

The proof of the pudding is in the tasting.

One version is "in the tasting" and another is "in the eating." This aphorism first appeared in literature in Cervantes' *Don Quixote*.

The idea is that it's not only what the pudding looks like, it's what the pudding tastes like—it's the doing, it's the action, it's the real world. It's not looking at the pudding.

Here is the real twist:

Usability = Can they do it?

Impact = Did they do it?

Can visitors to your Web site do something? Is the Web site designed in such a way that they can do what you want them to do? The real issue is, once they're free range chickens, can they do it? And the real test is, do they do it?

(Some) Outcome Assessment Methods

I'm now going to go through a lot of ideas very rapidly. I'll go into more detail about some of them than others, and the speakers who follow may talk about some of them more than I will.

Web Log Analysis

Most of you do Web log analysis now. Those of you who do it know how to do it better than me, and Rob Semper is going to talk about this during his presentation, so I'm not going to.

User Surveys

There are a variety of kinds of user surveys you can do to find out what the people who come to your site are doing, knowing, and leaving with. A lot of it has

(Some) Outcome Assessment Methods

- Web log analysis
- User surveys
- User focus groups
- Web usage diaries
- Analysis of user submissions and transcripts
- Institutional data analysis
- Assessments of learning
- Heuristic evaluation
- Off site actions

Web Log Analysis Things you can look for with Web Logs ...

- Number of Users
 - Total hits vs. unique IP addresses
 - Daily, weekly, monthly, etc.
- How long users stay online
- What pages/activities use or do
- What path do users take through the site
- What sites do users link from

User Surveys

Types of user surveys:

- E-mail invitation (e.g., site component analysis)
- Real-world user intercepts
- Pop-ups
- Self-selected surveys on the site

Things you can find with surveys:

- User demographics
- User attitudes or beliefs
- Related behaviors or experiences
- Users' wants and needs

to do with how you get them to be subjects. That's the real problem, isn't it? You have an effect, you know they're coming to your site. How do you reach out and grab them? One way is to invite them. Send out invitations that say, "We have selected you because we want people like you to come and view the site and answer some questions about it."

If you are a museum or science center, you can have the person at the gate hand out a piece of paper that says, "Come to our Web site." And there could be a special link on the site that says, "When you come to our Web site we want you to register just for this purpose, so we have a way of knowing who you were."

Pop-ups are increasingly important because nine out of ten times we gather some people from pop-ups. We can also put the pop-ups where we want to know something. There are a lot of people who have questionnaires or user survey forms on the home page. What is the last thing you do before you leave the site? You don't go to the home page. You exit the site from the last place you visited on that site, and if that's what we're interested in, that is where we want the survey to be.

You can have forty pop-ups in a rich site, and the same person is not going to get all forty of them. They may get two of them, they may not get any of them. In half-an-hour visitors can traverse a lot of territory and we can ask them a lot of questions about different things. Where is it that we want to know something? Where is it that there is an activity, where is it that there is a piece of knowledge that we want to know? That's where we want to put the pop-up.

The self-selected survey is what people are already doing now, and something like three to five percent of the people who come to your site are going to answer your survey. Some people get a higher return than that and if so, I bless them. They are really lucky.

What can you find out from surveys? You can find out about user demographics, you can deal with attitudes and beliefs, you can deal with the kinds of experiences they've had on the site, and you can deal with what they're thinking about doing after they're finished with the site. You can ask them a lot of different questions, and again, these are just samples. The real issue is, what do you want to know from somebody who is currently on your site?

User Focus Groups

Types of user focus groups:

- Face-to-Face
- Online/Virtual

websplorer jen:	did you, or would you play those games again?
Participant 1:	yes
Participant 2:	yes
Participant 3:	i like your in charge
websplorer jen:	did you find them challenging?
Participant 3 :	no
Participant 1 :	no
websplorer jen:	what could be done to make them more challenging?
Participant 1 :	make it a maze, or action game
websplorer melis:	is it important for a game to be challenging in order for you to like it?
Participant 1 :	no
Participant 2 :	nope
websplorer jen:	what kinds of things do you like in a game?
Participant 1 :	the action

There are two types of user focus groups: face-to-face and online. I've included an example above of a focus group with two moderators and about six kids. You need that many moderators because there are a lot of things that you need to prepare responses for. Some responses you can prepare ahead of time and just cut and paste them in when you need them.

The idea is that you can get a group online to do what is essentially a real focus group, a very traditional focus group, and it works fairly well. However, you have to keep kids on target and there has to be some control. That's why somebody deals with content while somebody else tries to manage the group. The kinds of questions you can cover with a user focus group can range anywhere from appeal and interest issues to "What would you name this?"

Let me give you some examples from some of the work that I've done over the years. I'm going to take this out of the museum context and into the commercial world and talk about a site for an OEM that does, among other things, digital cameras. They're interested in selling digital cameras, and they have a Web site on digital photography, but the Web site wasn't doing anything for the camera business. So we did a series of focus groups for them.

Among the things we did with the focus groups was show them different photo Web sites—Ofoto, Nikon, a whole range of things—and asked them some questions ("What if elements were here? What if elements were there?"). But we also said, "Here are fifteen cards, and on each is what could be a section of this Web site. Sort them into at least three and at most seven categories. Put a stick-on note on the top of each one and name them." We did that early in the

session and again later in the session.

What we found out was that people wanted names that meant something to them, and had no idea what some areas of the Web site were when they didn't understand the name for that area. As a result, they ended up changing the site dramatically. A focus group can help you understand what your client group really wants to know as opposed to what you think they should know.

I used to run pigeons and rats and had to watch their behavior (yes, I'm a behaviorist, but I don't know if I should admit that to this group). With the pigeons and rats, you had to watch them. With people, it really is interesting—you can ask them questions and, for the most part, they will tell you the truth. Many of them do lie—they want to make you happy and they lie about it—but the fact is, most of them will tell you something. So in many ways, using focus groups will give you a rich amount of information if you're willing to listen.

And it isn't just talking. Card sorts are one of the things you can do; you can put questionnaires in the middle of those groups; you can have them do activities of various kinds.

In the focus group on digital photography, one of the things that emerged is that people don't go to a digital photography site to find out about cameras. They want to find out about digital photography. They want to know how to do it and what's good and what's bad. We had people who were experts, mid-range users, and novices, and they were all interested in the same thing: How can I learn how to do this better? It turns out that a digital photography site is an education site. Most of the people go there

User focus group discussions can include:
(just a sample)

- Interest in concept
- Appeal
- Language issues
- Existing understanding of concepts
- Usability, functionality
- Play values, engagement, interactivity
- Current and potential audience, use patterns
- Marketing and promotion strategies
- Learning issues
- Connections to follow up activities

to learn, not to buy a camera. If they want to buy a camera they'll go to a bricks and mortar place where they can see the camera and touch the camera. Then they may go back online to find the cheapest price.

We also did something on digital entertainment, whatever the hell that is. An OEM wanted to do digital entertainment, but nobody knew what it was. It turns out that for early adopters it's really a geek thing. The OEM had all of this feel-good stuff about the benefits of hooking your television and music, and everything else together. What the people who felt they were in that field wanted to know was: "Let me see the plugs. I want to know what I can hook up to what." They didn't want to look at a beautiful picture and see lovely people reclining in their lounges, they wanted to see what hookups were there, what cables were needed, and how much the cables cost.

So the OEM had approached this in the wrong way, and they wouldn't have known that through questionnaires because people wouldn't respond to questionnaires like that. This is something that can emerge only in a discussion.

Web Usage Diaries

Web usage diaries are really interesting because they can give you some idea of how things are progressing over time, and that is one of the things that we rarely invest in. What we want is to be able to say, "Here it is, give me some feedback, and I'll change it." But what happens is that visitors do different things over time, even on the same site; even if it's rich enough to keep them coming back time and time again. For something like this, you really do need to recruit people and you need to give them an incen-

tive. Cash works. The idea is that if people are going to give you their time over a long period, you need to give them something in return, like a hundred dollars. In some ways that's a lot of money, but you don't need a lot of people.

What we've done in the past is send people daily e-mail questions or an e-mail with a choice: Press "respond" to answer this question, or use this link to a Web site and respond to the question(s) on the Web site. We prefer the latter because we can then synthesize their responses a bit more easily. Sometimes it's one question, sometimes it's a series of questions, and sometimes it's a focused set of questions: "We want you to take a look at this section of the Web site and respond to the issues that we're concerned about in this section. Don't look at anything else, just come."

What you're really looking at is how things change over time. What is of interest now that wasn't of interest when you first got on the Web site? The birding Web site (www.birdsource.org/gbbc/eBird.html) is a good example. I would want to know something about the people who keep coming back. What is it that interests them, and do their interests change over time? If, when they come back, they're now spending two minutes where they used to spend twenty minutes, what do we do to keep them there longer that would be useful to them? It's not about stickiness, it's really about how to increase the value that the site is providing.

Analysis of User Submissions and Transcripts

This is something that you all get—you get letters, you get responses from "Give us your feedback

Web Usage Diaries

- Recruited participants, with incentives
- Daily, weekly, and periodic questions
- Focused questions
- Looking for patterns, lasting impressions, unique experiences, factors that influence use

here,” and you get unsolicited information. You can really use that if you focus on how to collect it and how to organize it. Content analysis is a strategy that is real work, but it can help you.

Analysis of User Submissions and Transcripts

- Look at items that users have submitted or contributed
 - Content analysis
 - Analysis based on public rubric
 - Peer or expert critique
- Analyze transcripts from chats if they are available
 - Look for themes linked to objectives
 - Ideas for further site development

It can also help if you have some sort of rubric. Let’s take the *Backyard Jungle* Web site (www.pbskids.org/backyardjungle). Here’s a case where people have submitted information in the form of a picture or notes about their backyard. One thing you might want to look at is the kinds of backyards that young kids provide and how they’re different than the backyards that older kids provide. You might have a list of different types of things. For example, you look for vegetation of different kinds, animals of different kinds, backyards that include something that aren’t mammals, and so forth. You can play with this and figure out what kinds of things you want to stress or not stress. Should you put some information up there that says backyards can include things that are underground as well as things that are above ground? You begin to identify things that can generate more interest and more activity on the site.

Institutional Data Analysis

If you’re working in education, you’ve got a lot of information that other people are collecting for you and that includes the standard things that schools collect. And depending on what the educational focus of your site is and your links to formal education, there are different things that might be of interest to you.

If you’re in a different kind of institution, as many of you are who are at this conference, there are different kinds of things you might be able to deal with in your real site and your virtual site. I think Minda Borun is going to be talking more about this in her presentation. Many of the things listed below are things that you are doing anyway, but some of you might not be doing all of these things.

Institutional Data Analysis (Science Center / Museum)

- Bricks & Mortar
 - Changes in gate, membership
 - Store purchases linked to site content
 - Questions asked of staff
 - Traffic at specific exhibits
 - Enrollments, registrations in programs
- Web site
 - Downloads
 - Enrollments, memberships, purchases
 - Donations

At your real, bricks and mortar site, there are questions that are asked of staff and you can do something with that. What are the questions (beyond

Institutional Data Analysis (Education-focused)

- Test scores
- Attendance
- Number of Behavioral Referrals
- Course completion
- Job/college placement

Assessments of Learning

Web site testing

- Integrated assessments
- Linked/referred assessments
- Third party assessment
- Content, social, and procedural knowledge
- Transfer

directions to the rest rooms) that staff get asked and shouldn't those questions inform the exhibits? Ask staff to record the questions they are asked on note cards or on an inexpensive audio recorder; collect and analyze them. They might inform, not only the physical exhibit, but also the portrayal of that exhibit online.

There are also the enrollments and registrations for your evening or weekend programs. For example, in San Francisco the California Academy of Science has Saturday bird walks at eight in the morning. That's not my favorite time but, being the accompanying spouse, I went along with it. You have to call ahead of time and enroll for the bird walks and they are always oversubscribed. I don't know if the California Academy keeps track of how many people enroll for Saturday morning bird walks and how many people actually show. Sometimes people find out about the bird walks from a brochure, sometimes they may learn about them through the Web site. That's an indicator, in some ways. How do you use that information? They can use a questionnaire at the end of the walk with a smiley face that asks the question: "Did you find out about this from a brochure, from other people, or from the Web site?"

And again, there are a lot of kinds of institutional data that you can pull off of the Web site.

Assessments of Learning

In some cases you can integrate assessment into the Web site that you have. You can get people to take a test without them really thinking that it's a test. You can have people fill out materials that assess their knowledge and what they've learned before they

proceed with the game or to another area of your site: "We'd like to check in with you before you take the next step."

There are also referred assessments. A pop-up can say, "If you're willing, we'd like you to take a test." But you don't want to call it a test, so you say, "We'd like to know what you've learned so far." And you tell them that if they fill this out, you'll enter their name in a drawing to win a gift certificate, give them free entry to the museum, or offer them something else that you think they might want. You use that kind of incentive to get people to really engage, and many people will. Not everybody, but many will.

Then there is third party assessment. A teacher could assign your Web site as part of an educational activity. The teacher is, in fact, collecting data about your Web site. How do you win the teachers and set that up? You tell the teachers that you have materials that they can use in their classes and you would like to encourage them to do so by providing them with an educator's guide to the site. In return, you would like them to give you some information about what they assess in their classrooms to help you understand how useful or successful the Web site is.

You can also look at content, social, and procedural knowledge, and each of those may call for different strategies to get the information you want.

Transfer is another area. For example, if I'm learning about birds, do I then want to go out and look at birds? If I go to your site because I've seen a bird and I don't know what it is, and I'm going to your Web site to learn the name of the bird, its features, and its activities, do I then go out again afterwards to try

to find another example of that bird? And do I observe it in ways that I didn't before because now I know what I'm looking for?

Heuristic Evaluation

We've been doing some heuristic evaluation at this conference by having experts look at Web sites and respond to those Web sites. There are really two ways of using outside experts. One involves the content expert who asks, is this the right material? Does it satisfy the content needs of the audience? The other involves usability experts, like some of the Web people at this conference, looking at things such as navigation, the kinds of visuals you see on the screen, and so forth.

Heuristic Evaluation

What and Why?

- Content or Web Usability Experts
- Establish set of guiding questions or evaluation rubric
- Cheaper and quicker than full blown user testing
- Can help to focus future research efforts on most important elements

It is really helpful to bring in experts once in a while and have them look at your site in a particular way. You can charge them with the task of looking at particular issues that you are struggling with. The expert can focus on those things, they can give you some good information, and many will do it for free. They may help you identify where you need to make revisions, conduct more research, and do more work.

You can ask them to perform a task, just as you would a participant in a think-aloud, and see how they respond as experts.

Off Site Actions

Here is something that we have been talking about for the past couple of days: How do we get visitors to our site to do something we want them to do when they go off of the site? Are they going to go in the direction where our resources are leading them? There are a number of examples of off site actions that we could talk about, and this list is far from exhaustive. A lot depends on the individual issues. Longitudinal studies are interesting to me because nobody seems to be doing much in this area. We need to find out over time whether Web visitors have actually engaged in something related to our Web site that we encouraged them to go and do. We often say, "Okay, we've finished the site, and the funding is set up so that the day we're finished we're going to do the summative evaluation and that's it—we'll hand in the report, get our last check, and go for the next NSF proposal."

But the fact is that it doesn't happen overnight. You're called on to do things that may take you weeks. One of the things that we've done in the past with Web site evaluation is to first do a summative evaluation. Then, a month later, we will go back to about half of the people we talked to during the summative evaluation. So, for example, if we talked to one hundred people during the summative evaluation, we go to fifty of those people again after one month. In two months, we will follow up with twenty-five of those same people as well as twenty-five new ones. In three months, we will follow up

Off Site Actions (examples)

- Engage in activities
- Longitudinal studies
- Public participation
- Purchase decisions
- etc. (let's talk about them)

Issues to Consider

- Significant treatment (do you have one?)
- No silver bullet, no one solution for all
- No single strategy (triangulation)
- Timing (staged and iterative)
- External and/or internal
- Targeted vs. realized audience
- Critical competitors
- Budgeted activity

with twelve of those in addition to twenty-five new ones. We want to see if these visitors are continuing or if they begin to engage in things that they didn't in the first two days after the site was up.

We have been doing work in the San Francisco Bay Area with KQED, which includes a television station, a radio station, an education program, and a Web site. They are working on a new science project they've just submitted to NSF. They want the Web site to do original science material for people in the San Francisco Bay Area. One of the ideas they've come up with is to focus on science nature hikes. The example they use is that you could go on the Web site, see some hikes you could take and find out how long they are and how strenuous they are, and find out how to get there either by public transportation or driving. You could also find information about the history of the hiking site, learn about some science that you might see as you take the hike, and get information on games you can play or things you can talk about with children. The whole thing fits onto one page that you can print out and take with you as you head off for your hike.

One evaluation opportunity in this case is that when you go to a hiking destination there are often information sheets or maps at the trail head. How many people are taking the sheets? Once you put a Web site up and promote it, you can look for some of the outcomes you can measure that are associated with the activity that you've encouraged them to go out and do.

Issues to Consider

These are issues that I hope will generate some

discussion. First, people are looking for major outcomes when they have a minor treatment. You don't expect to change the world by having someone spend five minutes on your Web site. Get real. Figure out what is the order of magnitude of your treatment. If it is something to which people are going to keep coming back, that is a different level of treatment than if people are going to make a single visit, if they're going to go to your site and that's the last they're ever going to see of it. So think about it.

Second, there is no silver bullet. I can't say, "Here is a questionnaire that you should use." I can't say, "If you give kids who come to your Web site this test, you will have proved whatever."

You get something from this group, you get something from the same kind of visitors in a different fashion, and you get something from another, similar group in yet another fashion. You put it all together and you get a different view of the potential outcomes that you have. If the findings from all three data collection methods are consistent, then you can have confidence in the findings.

There is also the timing of all of this. You can stage assessment at multiple points over time, and it can be iterative in a way that allows you to learn more and more as you proceed.

Then there is the question of external or internal evaluation. There are some things that you can and should do yourself because there is information that you need to own. There are other cases in which you are in fact biased. You should recognize those cases and bring in people who don't have the same vested interest as those of you who created it. Sometimes you can do that cheaply, sometimes it's expensive. It

all depends on what you need.

Another issue is the targeted audience versus the realized audience. You have ideas about who you want to see coming to your site to do certain types of things, and you have often gone out and selected those people for testing. The fact is, you really need to look at both audiences. How can you develop strategies for bringing the people you originally wanted to get to your site? How can you take advantage of the people you actually got and move them in the direction you want?

Something we rarely do is take advantage of critical competitors. You need to know what your audience sees in other people's sites. It's worth asking the audience at your birding site, "What other sites do you visit to get information about birds? What is it

about their site that you like? What is it about our site that you like?" With that information in hand, can you compare the two sites and learn something from that comparison?

Finally, assessment has to be a budgeted activity in your institution or in your project. If it's an afterthought, it's not going to be worth anything to you, mainly because it isn't going to happen.

Resources

I've included some references that some of you may know. This is a small sampling. There are other references out there, and when my own Web site is revised (when I finish writing the intro pages), you will be able to find these articles as well as links to additional references on that Web site.

Resources

Children:

- *Usability of Websites for Children: 70 Design Guidelines*. Gilutz and Nielsen (2002)
- *Guidelines for Usability Testing with Children*. Hanna, Ridsen, and Alexander (1997)

General:

- *Research-Based Web Design & Usability Guidelines*. Sanjay J. Koyani, Robert W. Bailey, Janice R. Nall (<http://usability.gov/pdfs/guidelines.html>) (2004)
- *Paper Prototyping*. Carolyn Snyder (2003)
- *Observing the User Experience*. Mike Kuniavsky (2003)

Evaluating Museum Exhibits and Online Programs

Minda Borun, Museum Solutions;
 Director of Research and Evaluation,
 The Franklin Institute Science Museum

Evaluation and the Development Process

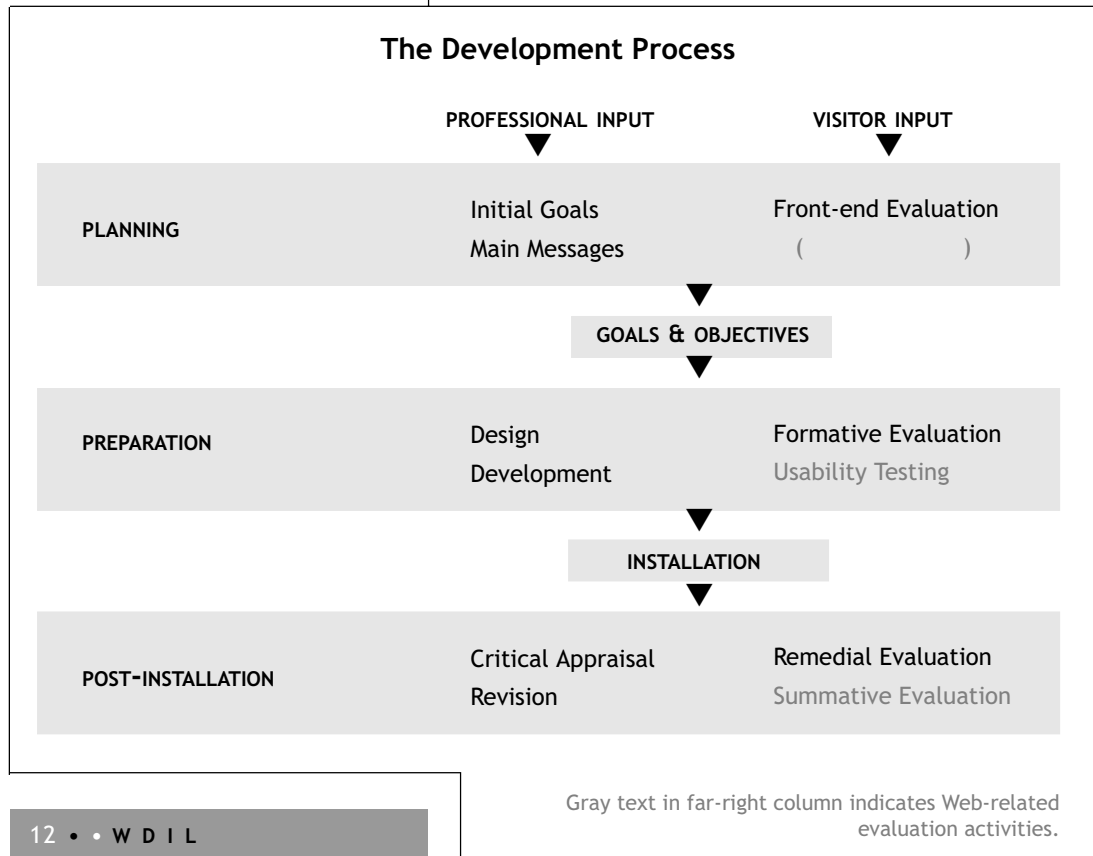
As most of you know, I work in a science museum and consult for museums of all sorts. My Web site experience is recent and somewhat limited. I've been

asked to talk about the differences between evaluating exhibits and evaluating Web sites.

The rather daunting diagram shown here is my way of illustrating that in the museum world, exhibit evaluation is not something that only happens at the end of the project. Evaluation is a process that parallels the development of exhibits. On the left side of the diagram you have the exhibit development process and on the right you have the corresponding evaluation activity. The black text indicates the procedures that we have for evaluating exhibits, and the gray text indicates the emerging procedures for evaluating Web sites and online activities. I've put parentheses under "front-end evaluation" for Web sites because it's my understanding that currently, there is little activity of this sort.

Front-end evaluation involves talking to members of the potential audience in the early planning stages of a project, before you've gone too far in development. This includes finding out what people know about the topic, what their expectations are, what misconceptions they may have, and which of various approaches might appeal to them and serve as a hook to get them to visit and stay on the site. I think there is a lot more work that can be done in this area, but there are also problems that I will get to shortly.

The second major bar in the diagram deals with formative evaluation—what Web folks call "usability and accessibility testing." Formative evaluation should be part of the ongoing evaluation process; not something that is only connected with the site development. It is a process that is repeated and should connect to summative or impact evaluation after launch.



Differences Between Exhibit and Online Program Evaluation

EXHIBIT	WEB
• Audience is known.	• Audience unknown.
• Goals defined.	• Goals broad.
• Visiting hours defined.	• Always open.
• Outcomes measurable.	• Outcomes difficult to measure.

There are differences between exhibit and online program evaluation. With exhibits, the audience is known, goals are defined, visiting hours are limited, and outcomes are to some extent measurable. With the Web, the audience is unknown (and I'll talk more about that), the goals are very broad and multipurpose, the site is always open, and the outcomes are difficult to measure.

Goals

EXHIBIT	WEB
• Goals apply to single exhibit or program.	• Goals are multiple and varied.
• Impact of whole visit measured in terms of visitor satisfaction (vs. learning).	• Goals tend to apply to a whole Web site (comparable to whole museum).

For exhibits, goals generally apply to a single exhibit or program. The impact of the whole museum visit is not specified and/or measured in terms of goals but

in terms of visitor satisfaction. With the Web, goals are multiple and varied and tend to apply to the whole Web site, which is comparable to the whole museum. Evaluation would be much simpler and clearer if the Web experience were broken down into components and you thought in terms of your success in achieving your goals for that component.

Audience

For exhibits, the audience is the local and tourist population with known or measurable characteristics. I can say this now, but ten years ago this was not the case. Ten years ago most museums had no clue who their visitors were because they had not yet asked them. Now, demographic surveys in museums are common and museums have a sense of who comes to visit. Also, there is a general pattern across museums that on weekdays you get school groups and on weekends you get casual visitors, including families and other groups. With the Web, there are two different kinds of audiences. There is the primary, or intended audience, which usually differs significantly from the secondary or actual audience—the people who come after launch. The secondary audience is very often global. This is not the case with “gated communities” or subscription audiences, where access is limited; in which case you are designing for a particular audience and that’s who comes. But if a site is free-access, the people who end up being the users may be very distant from the museum or client site. Also, I have found that the secondary audience often includes many teachers.

I’m going to tell a story to illustrate this point. I was working on an evaluation for a program that involved

EXHIBIT
• Local and tourist population with known or measurable characteristics.
• Demographics are available: <ul style="list-style-type: none"> - Weekdays-school groups; - Weekends-casual visitors (families and other groups).
WEB
• Primary (intended) audience differs from secondary (actual) audience.
• Secondary audience is global. May be very distant from museum or client site.
• Often includes many teachers.

aiphila.org/aie/page2.html

the Philadelphia Public Schools, the American Institute of Architecture, and WHYY Public Broadcasting in Philadelphia. It was a really interesting project that involved bringing architects into the classroom to develop projects on the built environment with kids and teachers. When the project was completed, a presentation on it was filmed by WHYY and became part of a show that aired on public television.

A Web site was created to allow kids and teachers to talk with the architects and with one another. The site also contained a description of the project, an area where teachers could talk to other teachers, and an area for uploading student work. It was a very rich site. It was intended to help the teachers work through the projects with the kids. When I did evaluation focus groups I found that none of the local teachers (the intended audience) were using the Web site. The reason for this was that, at that time, the computer was in the library, and when a teacher went to the library with the kids, the kids used the computers. Teachers couldn't go to the library during the workday because they were busy teaching the kids. A teacher could use a computer at home after hours, but teachers had enough homework to do and if there was a computer at home, they wanted to use it for their personal activities. So the local teachers did not use the Web site, but a lot of other people did. An online survey showed that the site was being used by teachers from all over the world and that they thought the site was a wonderful resource. Teachers in India and Africa were finding this site tremendously useful. But, it wasn't so for the intended audience.

Measuring Outcomes

EXHIBIT	WEB
<ul style="list-style-type: none">• Audience in limited area. Allows:<ul style="list-style-type: none">- Tracking and timing,- Unobtrusive observation,- Exit interviews.	<ul style="list-style-type: none">• Audience is readily available. Allows:<ul style="list-style-type: none">- Dwell time measures,- Tracking progress through site,- Exit interviews.

In terms of measuring outcomes, exhibits and Web sites have much in common. For an exhibit, the audience is in a limited area and this allows you to do tracking and timing studies, unobtrusive observations, and exit interviews. For the Web, the audience is readily available and this allows dwell-time measures, tracking progress through the site, and exit interviews. Below are some thoughts about methodological issues.

Methodological Strengths and Weaknesses

EXHIBIT	WEB
Strengths <ul style="list-style-type: none">• Can have face-to-face conversations.• Can observe exhibit use.	Strengths <ul style="list-style-type: none">• Can collect large samples quickly.• Can try multiple methods.• Can check server logs.
Weaknesses <ul style="list-style-type: none">• Process is time-consuming and labor-intensive.	Weaknesses <ul style="list-style-type: none">• Feedback is less precise

In exhibits, you can have face-to-face conversations, and you can observe exhibit use, but the labor involved is intensive and time-consuming.

For the Web, you can collect information from large samples of people quickly and try multiple methods. Also, you can check server logs and find out certain kinds of information from them, but the feedback is less precise. You don't always know to whom you're talking; they may not tell you what you want to know, and it's difficult to probe.

Questions to Consider

I'm going to leave you with a couple of questions to think about:

- What is the unit of assessment? What makes sense for your site? Is it the whole site? Is the site a game, or a single experience, or is it a complex, institutional Web site that has multiple components with different purposes that you might want to look at separately?

- Who is the audience? Is it the target audience, the actual audience, or both? If it's both, you have to evaluate at different times using different methods. For Web sites, evaluation doesn't end at launch. As the Web site continues, the user population continues to evolve and you need to tap into it periodically.
- Who defines desirable outcomes? Is it the Web site developer, the client, or both?
- Who requires evaluation? Is it just the funder? Are you doing evaluation because you have to? Or are you doing it because it's going to inform your process, make a better site, and help you keep the site current, active, and in touch with its audience?

Questions

- What is the unit of assessment?
- What is the audience—target or actual?
- Who defines desirable outcomes?
- Who requires evaluation?

Who's Out There and What Are They Doing Anyway?

A Personal Journey Through Metricland

Rob Semper, Executive Associate Director, Exploratorium

A lot of people are working in evaluation, and there are many evaluations underway, and we've heard this morning about some of them. My question is, do we really know who's out there, what they are doing, and how Web sites really fit into people's lives?

I would posit that we really have no idea who is at the other end of the wire. Except for the wonderful information from the log-in sites (and those of you who have people actually log-in are lucky—you actually get data), we really don't know very much about what's out there.

So I really want to talk a little about metrics and the question, how do we know and what should we expect in terms of metrics for our Web sites? I actually started thinking about this while doing a paper with Roland Jackson in 1998 called "Who's Out There" for the Museums and the Web conference. That was seven years ago, and I realize my journey has not gotten very far. I don't have very many answers to these questions even though I've been supported by a lot of really great people at the Exploratorium, including Sherry Hsi and Rob Rothfarb and others who have been working on this.

So who's out there? We talk a lot about how we find out what's going on. We have log analysis, on-line surveys, off-line surveys. I want to talk a little bit about how I got started in this. Below is a real Web log, as opposed to the blogs people talk about. This

is probably a millisecond snapshot off of our Web server of four different interactions. These are four different people doing different things on our Web site, and this is what a log looks like.



Web Log (A real blog)

I'm a physicist, I love to look at data, and this is an incredible amount of data. It is amazing that we have all of this data about people operating on our sites. Unlike any other medium we are working in, whether it is exhibits, or books, or television, we have actual data of everybody doing everything on our sites. And fortunately, there is software that exists—like Webtrends, or like Sawmill, which we use—that can take this data and actually make something of it.

The question is, what does this mean? I'm going to talk about two different examples of how we have been thinking about using this to try to understand our audience and some of the issues and problems we are confronting.

Who's out there?

- Log Analysis
- On-line Surveys
- Off-line Surveys

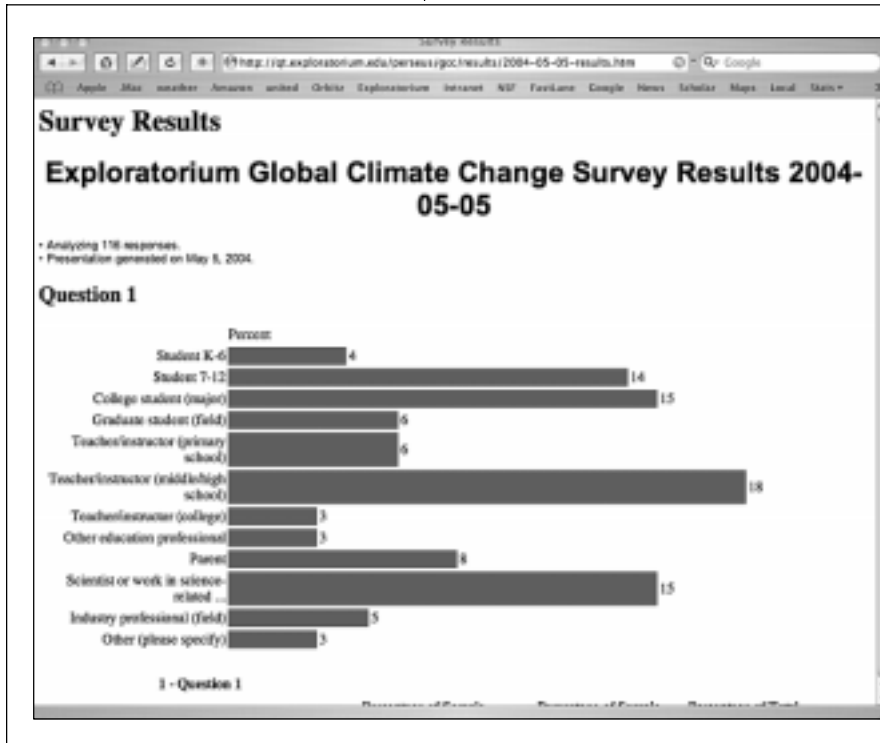
lot of visitors from K-12 sites. Therefore, we do know something: It appears that maybe more schools are going to the *Science of Music* site and more of the general public is going to the *Chaco Canyon* site.

On the other hand, we have considerable lack of knowledge here. We actually have a hard time knowing the individual nature of the people coming to our site. A lot of this is unresolved and we don't know much about it at all. That gets frustrating, so we decide to do surveys.

Below is another Web site, *Global Climate Change*

Research Explorer (exploratorium.edu/climate/atmosphere/index.html). With this one we moved up in the world. We used a Perseus Web site questionnaire survey and did some tests.

On the survey results page below, the bars represent various audiences, including K-6 students, college students, graduate students, scientists, and so forth. So you get all of this data, but what is interesting about this data is that it is all based on self-reporting. These are the people who really wanted to fill out the questionnaire. Does this represent our



The greenhouse effect is actually essential to our existence. The sun warms the earth, and certain gases (including carbon dioxide and water vapor) act like the glass of a greenhouse, trapping heat and keeping the planet's surface warm enough to support life. However, increasing humanity's effect on the concentration of greenhouse gases is a key issue in understanding global climate change. Industry and other human activity add carbon dioxide to the atmosphere. This strengthens the greenhouse effect and may cause a significant warming trend.

Understanding how the atmosphere works is fundamental to understanding climate change. The atmosphere is composed of layers of air, each with its own temperature patterns. Researchers must determine whether changes in temperature or air circulation are part of normal, long-term cycles, and the interconnections between air, sea, and land; and address that any change could have multiple causes—and multiple effects.

Global Climatic Increase: Heat Radiation - CERES Infrared Data
 short-term data

Global Reflected Shortwave Solar Radiation - CERES Infrared Data
 short-term data

Atmospheric Carbon Dioxide Records from Mauna Loa, Hawaii (1958 - 2002) - Carbon Dioxide Information Analysis Center
 short-term data

Global Average Mean Surface Temperature in Monthly Anomalies (1979 - 2001) - Global Carbon for Climate Processes and Research
 short-term data

Global Stratospheric and Tropospheric Temperature Anomalies (1979 - 2001) - Global Hydrology and Climate Center
 short-term data

glossary terms
 Click for definitions of words used on this page:

[carbon dioxide](#)
[climate](#)
[climate system](#)
[greenhouse effect](#)
[greenhouse gases](#)

more sites about the atmosphere
 • [NASA's Visible Earth: Atmosphere](#) - A directory of images, visualizations, and animations related to the earth's atmosphere.

sample? I don't know. There are people who have studied this, but I couldn't tell you whether or not it represents our sample.

It does point out the incredible variety of users coming to this Web site. And what does that mean? We're not actually in charge of who comes to our site. People discover the site. In some sense, we built a site that, for some reason, this diverse crowd of people are now using.

What are the metrics here? How do we measure this? If, for example, you wanted to do something for teachers and then found that all of these other people were visiting your site, what should be the appropriate measure for this?

- Who realistically is our audience?
- What are reasonable metrics of reach?

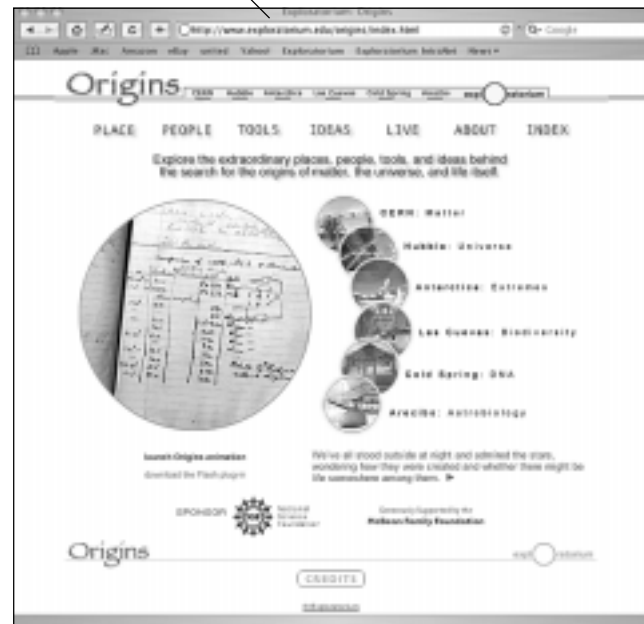
So who, realistically, is our audience? We are in a stew of opportunities. The audience is incredibly segmented. Given the *Google* world, what are the reasonable expectations for people coming to our site? Is it important to have one thousand people reached deeply or 100,000 people reached shallowly? Is the product of the two (audience x depth) actually the valuable measure that we should be using here?

Earlier at this conference, we heard a wonderful presentation on *Whyville* (whyville.net). Just for fun, I took the *Whyville* data and our data and, matching just reach of audience, I found an interesting thing. When they were up, we were down in terms of time

over a year, and when we were down, they were up. What does that mean? I have no idea what that means. Is it because kids are in a *Whyville*-type environment when they're not in our type of environment? Is it an artifact of bandwidth? I have no idea. That stuff is fascinating, and yet I don't think we know what is happening or why.

I'm going to move on to this other question—What are they doing anyway?—which is the other thing that this kind of analysis can tell us. There are a variety of ways that people are attempting to answer this question, including log analysis, tracking software, on-line surveys, focus groups, phone surveys, and ethnographic data.

Below, as an example of pushing on this, is the *Origins* project (exploratorium.edu/origins/index.html), where we actually did some tracking.



What are they doing anyway?

- Log analysis
- Tracking software
- On-line surveys
- Focus groups
- Phone surveys
- Ethnographic data

The *Origins* project was a Web site that involved going to different research locations around the world and showed the scientists doing their work *in situ*.

In this case, *Unwinding DNA* (exploratorium.edu/origins/coldspring/index.html), we talked with people at Cold Spring Harbor doing DNA research and genetic research.



Some interesting data came out of this that we communicated in our final report to NSF, which illustrates an interesting problem concerning metrics as well. Our data analysis shows that, on average, people spent four-and-a-half minutes on this site. They visited 2.3 pages on average, and ten percent came back for repeat visits.

Now, of course, the questions are coming, such as: How do we get people to stay longer? How can we

get more people to come back? I have a couple of responses to those questions. One, of course, would be to ask, what happens at exhibits? If you put an average marker on an exhibit, or even an exhibition, what would those numbers be like? What are the return numbers to our museum? It's weird in a way—we actually have more data here and in some sense, having more data is pointing out issues, and we have no metric to understand what that data means.

An even more serious consideration regarding this data is that, because of *Google*, probably a quarter of the people coming here are one-page wonders. *Google* sends people to this site because they're looking for "Cold Spring Harbor." Maybe they're just planning a visit to Cold Spring Harbor. They get to this page and their reaction is, I don't want this site, I want to go somewhere else. So one-page wonders completely skew this data because they contribute to that average of four-and-a-half minutes. If you scrape those people off, the average is probably more like ten, fifteen, or even twenty minutes. Maybe it's a half-hour, I don't know.

There was a physicist who visited the Exploratorium named Dennis Purcell, who used to do data analysis for ZDNet. He used to take their data and squeeze it through something called "Fourier transform," which is a physics process that turns time into space. In the process, he basically got rid of all of these one-page wonders and found all of this wonderful, enormously fascinating data. The point is, when you are reporting four-and-a-half minute average stay time on a site, is that rational or reasonable, and what is our metric? How do we know what our metric is? What is a usual metric for our work?

I also wanted to show you some of the things we're now playing with. We have gone in and put in some tracking software so that, at points when you are interested in a particular thing happening, you can actually get flags to understand what is happening.

Below is a *Math Explorer* site (exploratorium.edu/math_explorer/search), as well as an example of how we are tracking user activity. We're trying to understand how people make choices between activities and about where they go. So again, we are trying to get deeper data.

My point here is how do real users use our Web sites, and what is a good metric for good design? We know a lot about good design, but what really are our metrics?

- How do real users use our Web sites?
- What is a good metric of good design?



The screenshot shows a web browser window titled "Math Explorer Activity Database Tracking". The address bar shows the URL http://ip.exploratorium.edu/ipy/reporting/math_explorer/admin/index.php. The page displays a table of tracked activities.

Math Explorer Activity Database Tracking
 Connected to prodMathExplorer database on ip.exploratorium.edu
 There are 38 activities being tracked in the tracking table.
 IP filtering is ON

Activity Index	Activity Name	Click-through count
00000018	Bottle Blast-Off!	909
00000005	Fantastic Four	615
00000010	Mind Reader	593
00000020	Flying Things	511
00000001	Boxed In!	433
00000031	Build a Solar System	413
00000028	Dime Tic-Tac-Toe	399
00000013	Greeting Card Boxes	396
00000026	Jumping to Conclusions	359
00000017	Height Sight	357
00000003	Pig	324
00000012	Colossal Cartoons	301
00000025	Ladle Rat Rotten Hut	292
00000021	Stride Ruler	290
00000024	Centimeter Ruler	277
00000002	Oddball	270
00000014	Jacob's Ladder	260
00000007	Hopping Hundred	259
00000004	Madagascar Solitaire	258
00000000	Multiple Choice	177

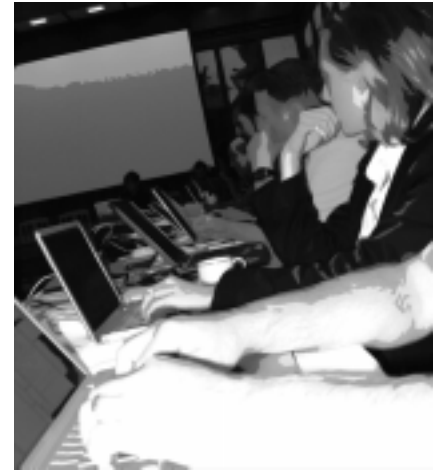


I will close by asking once more, who is out there and what are they doing anyway? What should our metrics of success be? How can we measure this? What does all of this have to do with learning?

I think we don't know very much, so we really have no idea.

Who's out there and what are they
doing anyway?

We still really have no idea.



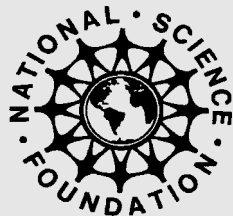
**Participate in this
Ongoing Discussion
Online**

Go to the Web Designs
for Interactive
Learning site
and join in this and
other discussions.

www.wdil.org

This document is intended to be a faithful synthesis of the discussions at the Web Designs for Interactive Learning conference that took place at the Cornell Lab of Ornithology in Ithaca, New York on June 15-18, 2005. It is meant to serve as a resource for those who attended and as a resource for others in the field. It does not necessarily reflect the views of the Cornell Lab of Ornithology, the Exploratorium, or individual conference participants.

In some sections, participant comments have been paraphrased. These are not exact quotes, rather they are an attempt to capture the content and meaning of the ideas presented.



This material is based upon work supported by the
National Science Foundation

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.



607.254.BIRD (2473) telephone
www.birds.cornell.edu

159 Sapsucker Woods Road
Ithaca, New York 14850



415.563.7337 telephone
415.561.0307 facsimile
www.exploratorium.edu

3601 Lyon Street
San Francisco, California 94123-1099