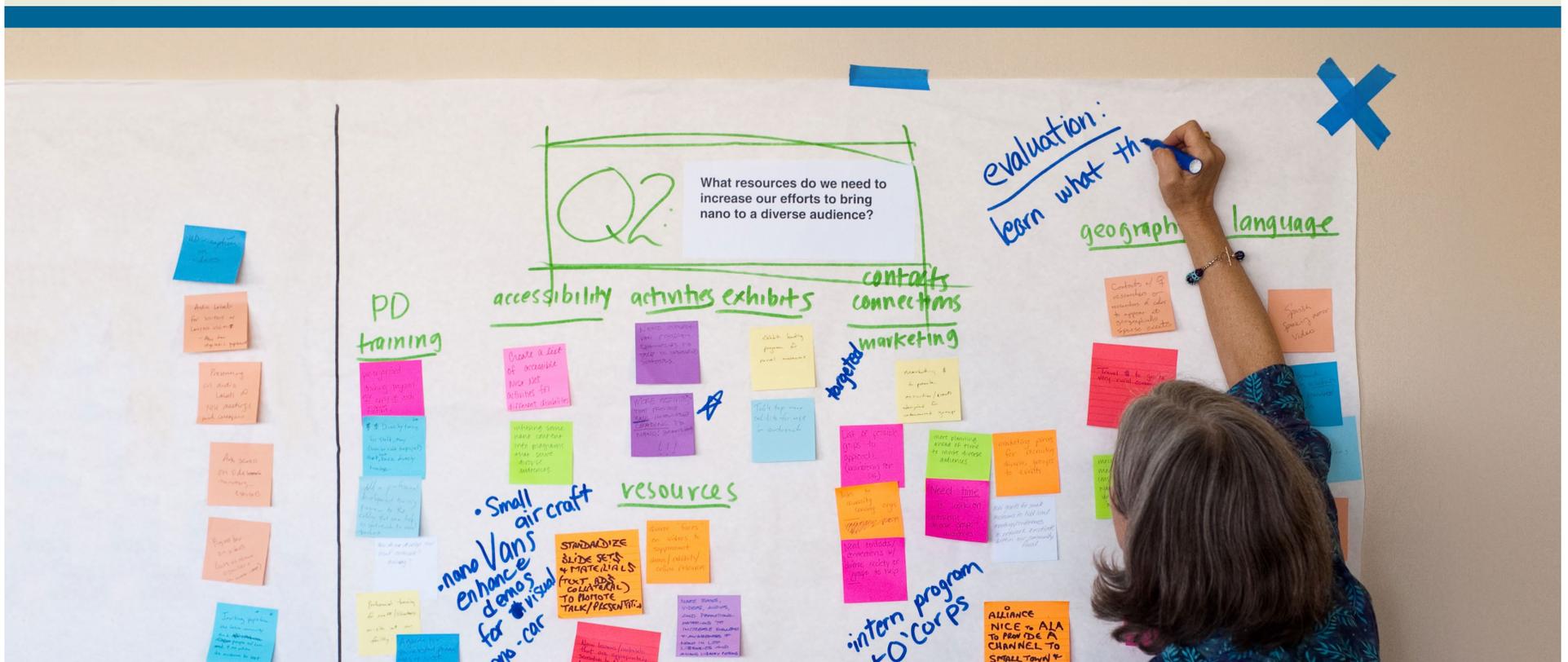


Evaluating the NISE Network

Reflections from the Evaluation Workgroup

Edited by Elizabeth Kunz Kollmann and Marta Beyer





www.nisenet.org

This report was based on work supported by the National Science Foundation under Award Numbers 0532536 and 0940143. Any opinions, findings, and conclusions or recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the Foundation.

*Copyright Science Museum of Minnesota
July, 2016*

Acknowledgments

The authors would like to thank the Network Executive Group and Network Organizing Group (NEGNOG), including Larry Bell, Paul Martin, Rob Semper, Brad Herring, Catherine McCarthy, Vrylena Olney, Rae Ostman, Eli Bossin, and Kayla Berry, for all of their help and support with this document, and for the evaluation over the 10 years of the NISE Net. We would like to thank all of the individuals who reviewed this document and provided us with feedback: Rei Cameron, Marley Steele-Inama, Ali Jackson, Kathayoon Khalil, and Amanda Krantz.

Finally, we would like to acknowledge the dedicated efforts of all of the evaluators and Committee of Visitors members who have been involved in the NISE Network Evaluation since it began, including the following individuals:

Museum of Science

- Emily Anstey
- Ryan Auster
- Marta Beyer
- Kerry Bronnenkant
- Clara Cahill
- Elissa Chin
- Sunewan Chunhasuwan
- Amanda Colligan
- Sarah Cotterill
- Juli Goss
- Joseph Johnson
- Elizabeth Kunz Kollmann
- Anna Lindgren-Streicher
- Alexander Lussenhop
- Sarah May
- Iyar Mazar
- Leigh Ann Mesiti
- Jane Morgan Alexander
- Christine Reich
- Katie Todd

Science Museum of Minnesota

- Marjorie Bequette
- Sarah Cohn
- Patrick Cox
- Kirsten Ellenbogen
- Amy Grack Nelson
- Steve Guberman
- Lura Harvey
- Denise Hanh Huynh
- Gretchen Haupt
- Beth Janetski
- Zdanna King
- Eric LaPlant
- Elizabeth LaPorte
- David Milavetz
- Jane Miller
- Kathleen Miller
- Al Onkka
- David Ordos
- Gayra Ostgaard
- Claire Phillipe

- Molly Phipps
- Murphy Pizza
- Beth Robelia
- Maggie Ryan Sandford
- Amanda Svantesson-DeGidio
- Scott Van Cleave
- Levi Weinhagen

Oregon Museum of Science and Industry

- Marcie Benne
- Gabby Burlacu
- Chris Cardiel
- Scott Ewing
- Melanie Francisco
- Kari Jensen
- Brett Kiser
- Melissa Laurie
- Jenna Lecomte-Hinely
- Maria Montiel
- Scott Pattison
- Smirla Ramos-Montañez
- Scott Randol
- Nelda Reyes
- Liz Rosino Wright
- Debbie Siegel
- Anne Sinkey
- Mary Soots
- Omar Vargas
- Barry Walther

University of Notre Dame

- Gina Svarovsky

Multimedia Research

- Barbara Flagg
- Valerie Knight-Williams

Inverness Research Associates

- Pam Castori
- Jennifer Helms
- Mark St. John
- Michelle Phillips
- Dawn Robles

Exploratorium

- Veronica Garcia-Luis
- Josh Gutwill
- Sherry Hsi
- Adam Klinger
- Joyce Ma

Committee of Visitors

- Frances Lawrenz, University of Minnesota
- Bruce Lewenstein, Cornell University
- Saul Rockman, Rockman et al
- Carol Weiss (deceased), Harvard University

Contents

Introduction	5
CHAPTER 1:	
Setting Up an Evaluation Workgroup for a Network	
<i>Sarah Cohn and Elizabeth Kunz Kollmann</i>	12
CHAPTER 2:	
Measuring Network Impacts	
<i>Marcie Benne and Marjorie Bequette</i>	25
CHAPTER 3:	
Adapting Methods for Network Use	
<i>Sarah Cohn, Juli Goss, Liz Rosino Wright, and Gina Svarovsky</i>	36
CHAPTER 4:	
Creating Ethical Evaluation Practices for a Large Network	
<i>Elizabeth Kunz Kollmann</i>	46
CHAPTER 5:	
Communicating Findings and Methods to Network Stakeholders and Beyond	
<i>Marta Beyer</i>	56
Conclusion	66
Glossary of Common Terms.....	69
Notes.....	70
Appendix A: Content Map	71
Appendix B: NISE Network Goals for Professional Development	72
Appendix C: NISE Network Years 6–10 Logic Model	73

Introduction

Elizabeth Kunz Kollmann

Over the past 20 years, informal learning environments have evolved rapidly. One of the biggest trends in informal science education is the uptick in the number of networks and multisite projects as a way of sharing educational products and professional development more broadly. Thinking about evaluating these distributed and complex projects can be daunting. Where do you start? What questions do you ask? What data do you collect?

The Nanoscale Informal Science Education Network (NISE Network) Evaluation workgroup grappled with these questions for 10 years. Although we do not claim to have all the answers, we think it could be helpful for you to read about where we've been, the challenges we've faced, and the lessons we've learned. So, we created this reflection document to describe how we managed the evaluation of the NISE Network, one of the largest informal education networks ever formed. This reflection document is not meant to be a prescriptive “how-to” manual, but an example of one approach to the messy work of network-wide and multisite evaluation.

Who might benefit from this document?

This reflection document is specifically intended for colleagues who are involved in evaluating, or are planning to evaluate, informal education networks and multisite projects. In addition, educators and other practitioners working on these kinds of projects may find this reflection document useful when engaging in conversations about evaluation with project evaluators or project leadership.

Because many factors play a role in deciding the kind of evaluation you pursue, we recognize that your evaluation may not look anything like the evaluation we conducted for the NISE Network. We offer our team's experiences and decisions for you to consider while managing your evaluation project. For example, we've found in planning and implementing the NISE Network evaluation that the following unique characteristics of the NISE Network impacted the decisions that we made:

- **Size:** From the beginning of the Network, products were being created and implemented at multiple institutions. By the final years of the NISE Network, over a dozen institutions were creating products, professional development was being conducted for thousands of informal science education (ISE) professionals, and exhibits and activities were being used at hundreds of ISE organizations. The Evaluation workgroup had to determine how to evaluate a project that was spread across the country and had thousands of participants, both public and professional.
- **Timeline:** Not only was the size of the NISE Network large, but the amount of time that the project lasted was longer than many ISE projects. Over the 10 years of the NISE Network, the roles of partners changed, the roles of the workgroups changed, and, at

some points, the goals for the project and its deliverables evolved. This meant that the Evaluation workgroup had to be responsive and modify its work to reflect the shifting needs and desires of the NISE Network.

- **Content:** Nanoscale science, engineering, and technology (nano) had not been widely presented in ISE institutions prior to the NISE Network, and ISE professionals were generally not familiar with this topic or how to present it to the public. Therefore, the NISE Network had to learn how to best engage the public in nano content through this project, and the Evaluation workgroup had to come up with processes and methods to evaluate products about a new and emerging science, technology, engineering, and mathematics (STEM) field.
- **Open-source philosophy:** It was important to NISE Network Leadership that partner organizations felt free to modify and adapt Network products so that they would work within partners' individual organizations. The Evaluation workgroup had to figure out how to generate valid findings about the impact of specific deliverables that were not necessarily presented in the same way across sites.

These issues, and more, are likely to impact the evaluations of other network and multisite projects as well. We hope you will benefit from hearing how we approached the various evaluation challenges of the NISE Network.



How should I use this document?

This reflection document is organized around some of the main issues that we encountered in managing the evaluation of a large-scale informal science education network and the lessons that we learned in navigating these issues. We do not expect you to read the entire document. If you have a specific question about managing your project's evaluation, you can jump to the chapter that focuses on that topic. Our chapters cover the following:

- Chapter 1: Setting Up an Evaluation Workgroup for a Network
- Chapter 2: Measuring Network Impacts
- Chapter 3: Adapting Methods for Network Use
- Chapter 4: Creating Ethical Evaluation Practices for a Large Network
- Chapter 5: Communicating Findings and Methods to Network Stakeholders and Beyond

To help make the chapters stand alone, each one includes a similar structure. The first page of each chapter includes a description of the challenge that is being addressed and a summary of what is discussed. Inside each chapter, you will find contextual information about the situation that we faced, a description of how we managed this issue within the NISE Network, and some things to consider when you are dealing with a similar circumstance. Additionally, each chapter includes a short vignette that describes a struggle we encountered and how we resolved it.

We sometimes use terminology in this document related to the NISE Network or evaluation in general that may not be familiar to

everyone or may have specific NISE Network meanings. To ensure that you understand the way we use these terms throughout the document, there is a glossary at the end of this document (page 69).

What was the NISE Network?

The NISE Network was a national community of researchers and informal science educators dedicated to fostering public awareness, engagement, and understanding of nanoscale science, engineering, and technology.

Launched in 2005, the Network, as of 2015, included over 500 museums, universities, and other organizations working together to engage public audiences in nano. There were over 200 NISE Network-developed programs, activities, exhibits, and media free to download and use from its online digital library (www.nisenet.org) and more than 50 professional development and training resources. NanoDays, the NISE Network's signature event, annually mobilized hundreds of Network



partner organizations across the country to engage staff, volunteers, and members of the public in learning about nano. By the end of 2015, the final full year of funding, we estimated that 30 million people participated in NISE Network programs, events, and exhibitions.¹

The Network was funded by the National Science Foundation (NSF) through two successive awards of over \$20 million each: the first in 2005, and a renewal in 2010 (DRL- 0532536 and DRL-0940143). The scale and scope of the awards were unlike most NSF grants for informal science education. At \$41 million over the 10 years, the awards were substantially larger and the span of the project longer than is typical. Through the NISE Network, NSF aimed to create a field-wide infrastructure. The funding came not just from the informal science education budget but also from the following NSF directorates: Education and Human Resources; Biological Sciences; Computer and Information Science and Engineering; Engineering; Geosciences; International Science and Engineering; Mathematical and Physical Sciences; and Social, Behavioral, and Economic Sciences.

The scope of the work was enormous. Over the years, the Network encompassed over 20 different workgroups, each made up of staff from nearly 20 organizations, operating with varying levels of coordination and interconnectedness. The Network created a myriad of products, trainings, and dissemination materials, ranging from large and small exhibits to research studies on aspects of the Network. Developing shared understandings of goals and strategies, communicating across the groups, and sharing findings and lessons learned were all significant challenges. And, as the Network matured, its structure and its evaluation questions continued to evolve.

The Network was built on the belief that one hundred organizations would participate as partners. When this was achieved, the Network

set its sights on reaching hundreds of partners. The Network understood that each organization would bring their own diverse goals, motivations, and audiences, and fit their involvement to their own unique circumstances. Network partners were, indeed, diverse. There were rural and urban institutions; small museums with fewer than 10,000 annual visitors and large museums with over 750,000 visitors per year; and also different types including science museums, children’s museums, universities, and libraries. The Network was designed such that it would be able to function at this scale and in these profoundly varied circumstances.



Who was the NISE Network Evaluation workgroup, and what did we study?

Since the inception of the NISE Network, six internal and external evaluation groups, consisting of over 30 individual evaluators, participated in the Evaluation. In Year 4 of the NISE Network, the

Evaluation workgroup was transformed into a partnership among evaluators at three organizations, the members of which are the authors of this document. This multi-institutional Evaluation workgroup was composed of over 20 evaluators from research and evaluation departments at the following three science museums:²

- Museum of Science, Boston (MOS)
- Oregon Museum of Science and Industry (OMSI)
- Science Museum of Minnesota (SMM)

We were overseen by an external Committee of Visitors (COV), which is described in more detail in Chapter 1. Our COV was composed of Frances Lawrenz, Ph.D., an evaluation scholar from the University of Minnesota; Bruce Lewenstein, Ph.D., a science communication scholar from Cornell University; and Saul Rockman, a professional evaluator specializing in informal science education from Rockman et al. A former member of our COV is the late Carol Weiss, Ph.D., an evaluation scholar from Harvard University.

As evaluators for the NISE Network, we were tasked with conducting evaluations to both inform the Network's directions and understand its impacts. Therefore, the Evaluation workgroup conducted or mentored groups through front-end, formative, and summative evaluations. Our work focused on two audiences: the professionals who implement NISE Network products and the public who interact with NISE Network products. We will talk more about how we measured the NISE Network impacts and some of the methods we used in upcoming chapters.

In the first years of the Network, there were some common messages, especially for public deliverables. However, there were also some



goals unique to each deliverable, making it difficult to evaluate the impact on audiences across them.³ In part because of this, as well as due to the evolution of the NISE Net, Network Leadership crafted a revised set of common goals and messages for professionals and the public around Year 4. The resulting documents included:

- A content map that described key science concepts for engaging the public in nanoscale science, engineering, and technology (Appendix A);
- A complementary learning framework that describes the kinds of learning experiences the Network values; and
- Professional development goals that covered what the NISE Network wanted professionals to achieve through programs, workshops, and guides created for ISE and university professionals (Appendix B).

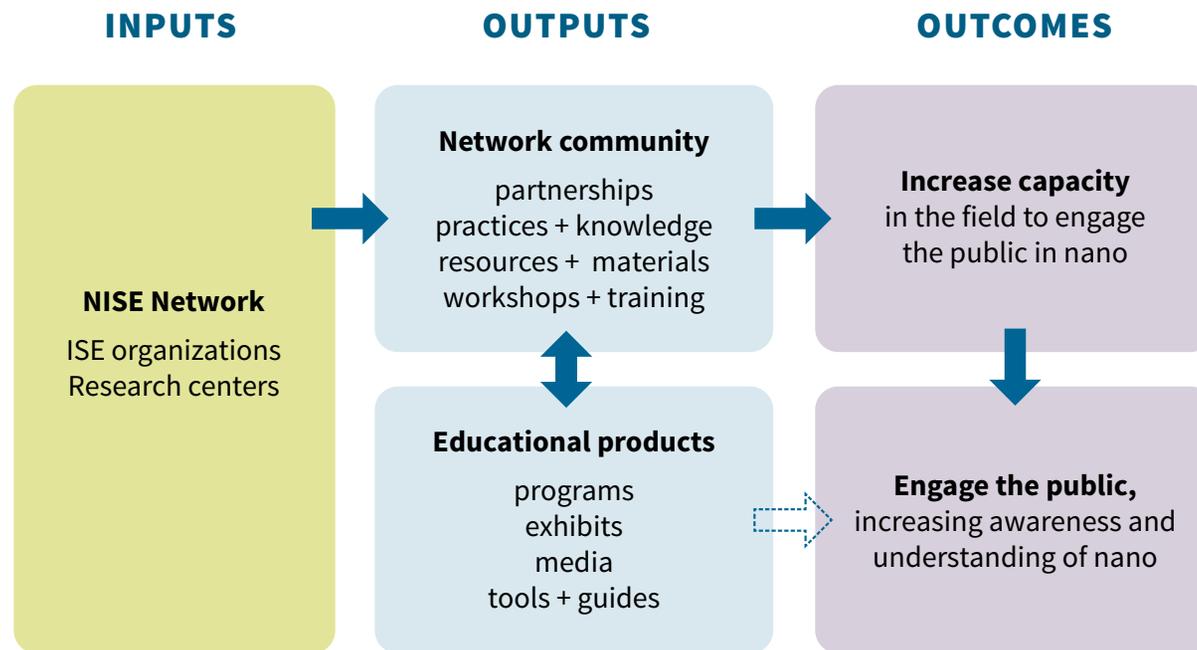


Figure 1. Simplified NISE Network logic model

At around the same time, the Evaluation team and Network Leadership worked together to create a detailed logic model to describe the theory for how these goals were going to be achieved (see Figure 1 above and Appendix C). Because the logic model laid out the process for how the NISE Network would achieve its impacts, the Evaluation workgroup began to use the logic model to plan their summative evaluations beginning in Year 6. Additionally, around Year 6, formative evaluation continued as workgroups created new

products. However, to deal with an increasing demand for formative evaluation, we moved from a model where evaluators conducted these evaluations to a Team-Based Inquiry (TBI) model where evaluators supported NISE Network partners in conducting their own formative evaluations as described in Chapter 3.

As the NISE Network began to wind down, it became increasingly important to make sure that we were capturing the major outputs and outcomes of the Network that had not yet been documented

through previous evaluation work. To achieve this, the Evaluation workgroup formed subgroups to ensure that we would have all of the data that we needed to describe the cumulative impact of the NISE Network. These teams were composed of a lead organization (MOS, OMSI, or SMM), which provided leadership and staff support, and thinking partners from the remaining organizations. This team structure is described in more detail in Chapter 1. These subgroups remained in place for NISE Network Years 8–10, and they covered the following aspects of the Network:

- **The Logic Model Group** was created to monitor the Network’s logic model to ensure that the different evaluation studies reflected the Network’s current thinking as it evolved over time.
- **The Public Impacts Team** investigated the impact of NISE Network on the public, including reach and learning by studying the Network’s biggest public deliverables: the *Nano* mini-exhibition and NanoDays.
- **The Professional Impacts Team** explored NISE Network’s impact on its professional audience (ISE and university partners) during the last three years of the Network in terms of effects on professionals’ sense of community, learning about nano, and use of nano education products and practices.
- **The Survey & Data Mining Team** coordinated and implemented an Annual Partner Survey and catalogued and mined various sources of data in order to inform other evaluation studies and workgroups.

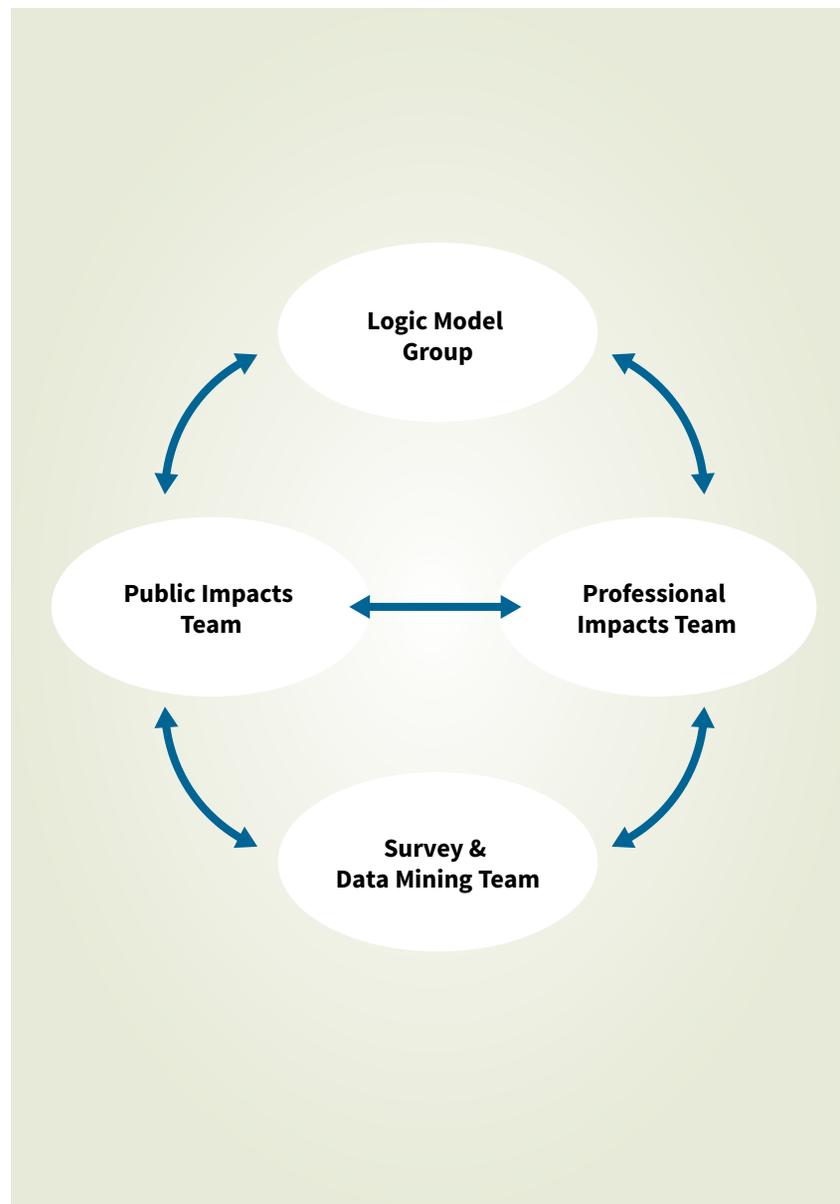


Figure 2. Evaluation workgroup structure for Years 8–10

CHAPTER 1:

Setting Up an Evaluation Workgroup for a Network

Sarah Cohn and Elizabeth Kunz Kollmann

The Challenge: How can we create an objective Evaluation workgroup for a large and changing network that is responsive to the Network's needs?

Chapter Summary: Over the 10 years of the NISE Network, the Evaluation workgroup was structured in a number of ways to fulfill the Network's evaluation needs. Important to the success of each of these structures was constant communication among the different Evaluation groups and with Network Leadership, as well as being embedded with other workgroups throughout the Network. This chapter describes:

- How the Evaluation workgroup changed structure over time;
- How we managed to create a cohesive, multi-institutional evaluation; and
- How we worked to remain flexible and responsive to Network needs while also being objective.

What we did

Evaluation was an important part of the NISE Network from the very beginning. From the first National Science Foundation (NSF) call for proposals in 2005, the need for evaluation to inform project directions and to understand the impacts of the Network was paramount for the project. Though the evaluators involved in the Network and the structure of the Evaluation workgroup changed over time, this directive has always been the focus of our work. To fulfill this directive, we had to manage a number of factors, including:

- How to produce a cohesive evaluation across a multi-institutional Evaluation workgroup;
- How to generate an objective evaluation; and
- How to remain responsive and flexible in order to fulfill the changing evaluation needs of the Network.

This chapter describes the history of the NISE Network Evaluation workgroup, recounting the various Evaluation workgroup structures

over the years, specifically sharing what worked well within a particular setup, and what required more attention or consideration. Throughout these descriptions, it is evident that managing the evaluation within the NISE Network environment worked best when the Evaluation workgroup members were embedded in the NISE Network workgroups for which they were conducting evaluations, and when we continuously communicated with each other. The chapter is split according to the three main eras of the Evaluation workgroup:

- **Years 1–3:** Using internal formative evaluators and external summative evaluators
- **Years 4–5:** Transitioning to an all-internal Evaluation workgroup
- **Years 6–10:** Having internal evaluators conduct summative evaluation and practitioners conduct formative testing with evaluation mentors

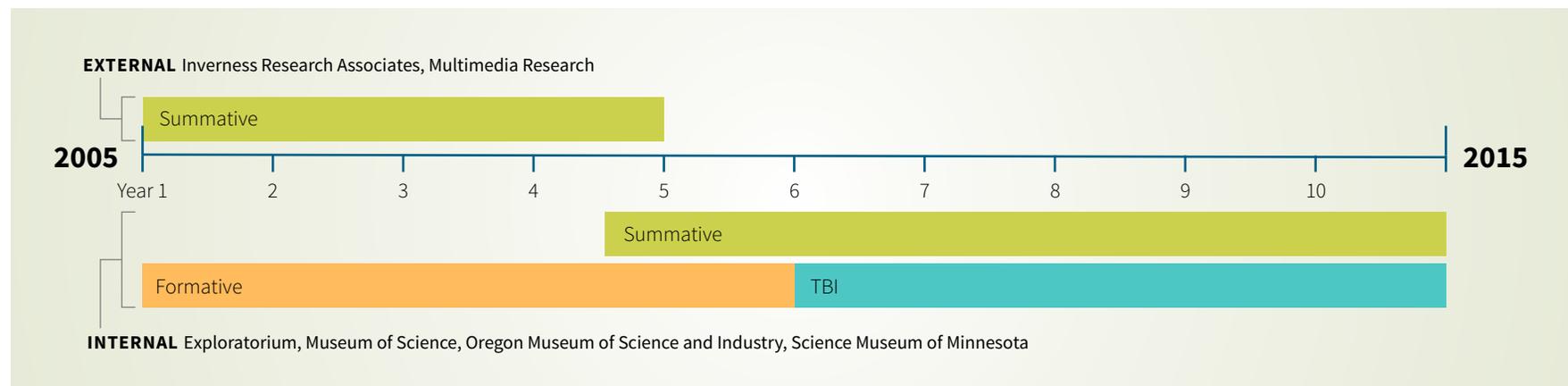


Figure 3. Evaluation timeline

YEARS 1–3:

Using internal formative evaluators and external summative evaluators

Due to the scope and scale of the NISE Network, multiple institutions have been involved in evaluating the Network since the beginning of the project. For the first three years of the Network, evaluation work was split in a way that was then the norm for NSF, our funder: internal evaluators conducted the formative evaluation and external evaluators conducted the summative evaluation. In matching with their expertise, Inverness Research Associates was chosen to conduct a summative evaluation focused on understanding the growth and development of the Network as well as professional impacts, while Multimedia Research was chosen to conduct summative evaluations focused on understanding the NISE Network public impacts. Internal evaluators from the Exploratorium, Museum of Science (MOS), and Science Museum of Minnesota (SMM) were embedded within the different NISE Network workgroups as formative evaluators, attending meetings just as if they were team members. These internal evaluation institutions were chosen, in part, because they led the NISE Network project, and therefore, much of the development work would be conducted at their institutions.

Workgroups during Years 1 and 2 included Network Administration, Annual Meeting Planning, Exhibits, Forums, Nanoscale Education Outreach, Network Media, Programs, Visualization Lab, and Website. Workgroups begun between Years 3–5 included Content Steering, Community, Inclusive Audiences, NanoDays, and Research Center - Informal Science Educator Partnerships (RISE). For each of these workgroups, there was a team composed of members from multiple institutions. At least one formative evaluator was a member of

each of these teams so that they could collect data as needed.

Summative evaluators worked with these groups to conduct studies as deliverables neared completion.

Years 1–3 Further details about how the team was structured during this phase to deal with cohesiveness, objectivity, and flexibility can be found below along with information about how well this worked within the NISE Network:

MANAGEMENT CONSIDERATION	WHAT WE DID	WHAT WORKED	EVALUATION CHALLENGES
<p><i>Creating a cohesive evaluation across a multi-institution team</i></p>	<ul style="list-style-type: none"> • Work was split between phases of evaluation with the external team taking on front-end and summative work and the internal team taking on formative work. • External evaluators split the work between impacts related to professionals/the Network and those related to the public. • Internal evaluators split the work according to the different workgroups with each one having their own evaluator. • Leaders of the internal and external evaluation groups met, as needed, to discuss work. 	<ul style="list-style-type: none"> • By splitting work by strand, phase, and audience type, there was a clear delineation of work. • Meetings between evaluators allowed them to learn new evaluation techniques from each other and use them across NISE Network studies. 	<p>Some of the educational products had different goals and formats, and evaluators did not meet regularly to discuss every aspect of our work, making it difficult to ensure consistency; subsequently, each workgroup evaluation used separate data collection instruments and methods.</p>
<p><i>Generating an objective evaluation</i></p>	<ul style="list-style-type: none"> • External evaluators joined workgroups and completed summative evaluations as needed when products and practices were completed. • Internal evaluators were embedded in workgroups and completed formative evaluations throughout the development process as needed. 	<p>By having evaluators who were both continuously immersed and less frequently involved in the workgroups, both a field-wide and network-specific viewpoint were represented.</p>	<p>Because of the ways that the different evaluators interacted with the NISE Network workgroups, formative evaluations often presented findings without a field-wide view, and summative evaluations did not always represent the nuances of the NISE Network deliverables.</p>
<p><i>Remaining responsive and reflexive to the changing needs of the Network</i></p>	<ul style="list-style-type: none"> • Internal evaluators attended all work group meetings. • External evaluators worked with Network Leadership and teams to craft evaluations, meeting with them as appropriate. • Reports were created to share findings with NISE Network partners, leadership, and funders. 	<p>By varying the level of evaluator involvement in the workgroups, formative evaluators could ensure that early rounds of data were collected at a time and in a way that made sense for each work group while summative evaluators could wait to become involved until products were nearly complete, saving time and resources.</p>	<p>Summative evaluators were less able to make quick adjustments in order to respond to the Network’s changing directions due to distance, scope, and time constraints, while formative evaluators had to ensure that all deliverables were neither over- nor under-evaluated.</p>

YEARS 4–5:

Transitioning to an all-internal Evaluation workgroup

Years 4 and 5 were a time of change for the NISE Network Evaluation. The summative evaluation work that external evaluators Inverness Research Associates and Multimedia Research had been contracted to complete was wrapping up, but additional formative and summative evaluation work needed to continue. At the same time, the NISE Network was planning to submit a proposal renewal for five more years of work. Therefore, decisions needed to be made about who was going to conduct different pieces of evaluation work and how. NISE Network Leadership wanted an evaluation that could (1) be responsive to their needs, (2) provide information that would inform their decision-making in a timely manner, and (3) inform the Network, the informal science education (ISE) field, and NSF about the project's impacts. The NISE Network Leadership group considered their options in consultation with current Evaluation workgroup members, and Network Leadership decided that the formative and summative evaluation needs would be integrated. This decision led to an Evaluation workgroup composed of internal evaluators from three institutions: MOS, SMM, and OMSI.⁴ Network Leadership thought that an internal team would have more flexibility to be embedded in and, therefore, respond to workgroups, and also that they would have the capacity to share evaluation findings more quickly. This structure was seen as optimal for the continued evaluation work.

As we moved from a split structure of external and internal evaluators, we wanted to make sure that we (1) did not lose the field-wide perspective that the external evaluators provided, (2) remained objective, even when we were embedded within workgroups and

NISE Network partner institutions, and (3) could offer support to the Evaluation workgroup's professional development and growth. To achieve these goals, the NISE Network created a Committee of Visitors (COV) that would provide an external view to the evaluation work. We chose COV members with diverse expertise that was not available among the internal evaluators, who would bring in new perspectives and knowledge to cause the Evaluation workgroup to think deeply about our methods, and who would help us grow as professional evaluators.

Years 4–5 Further details about how the team was structured during this phase to deal with cohesiveness, objectivity, and flexibility can be found below along with information about how well this worked within the NISE Network:

MANAGEMENT CONSIDERATION	WHAT WE DID	WHAT WORKED	EVALUATION CHALLENGES
<p><i>Creating a cohesive evaluation across a multi-institution team</i></p>	<ul style="list-style-type: none"> • External evaluators from Inverness Research Associates completed summative evaluations about professional impacts and network development. • External evaluators from Multimedia Research completed summative evaluations about public impacts. • Internal evaluators took on the remaining formative and summative evaluation work. • Internal evaluators worked closely with each other to determine how to split the new summative evaluation work related to public impacts as external evaluators completed professional impacts work. • Leaders of the SMM, MOS, and OMSI evaluation departments met at least once a month to discuss work. 	<ul style="list-style-type: none"> • Internal evaluators could use what they knew from the formative work to craft and conduct summative evaluations. • Because meetings between evaluators happened more regularly (at least once a month), they were better able to share their experiences and expertise. 	<p>The increased need for formative evaluation, in conjunction with the new summative evaluation work, outstripped the capacity and budget of the internal evaluators.</p>
<p><i>Generating an objective evaluation</i></p>	<ul style="list-style-type: none"> • A COV was added to provide additional expertise, advice, and oversight to summative evaluations being conducted by internal evaluators. • Internal evaluators used multiple methods for summative evaluation studies. 	<ul style="list-style-type: none"> • The COV provided a field-wide view of the evaluation as well as additional expertise that was lost when the external evaluators left the project. • Use of multiple methods allowed evaluators to ensure valid findings. 	<p>Evaluators had to balance the competing interests of the NISE Network and the broader field by figuring out how to provide information to inform decision-making in the short-term and describe project impacts in the long-term.</p>
<p><i>Remaining responsive and reflexive to the changing needs of the Network</i></p>	<ul style="list-style-type: none"> • External and internal summative evaluators worked with Network Leadership and workgroups to craft evaluations. • The leader of the Evaluation workgroup attended weekly Network Leadership calls to provide evaluation findings as needed and hear about Network plans. 	<p>Because internal evaluators left some space for flexibility and change in evaluation plans, they could ensure evaluations would be useful to the current needs of NISE Network Leadership and workgroups.</p>	<p>Sometimes the Network’s evaluation needs came in waves that caused a backlog, making it difficult for both internal and external evaluators to provide evaluation findings quickly.</p>

YEARS 6–10:

Having internal evaluators conduct summative evaluations and practitioners conduct formative testing with evaluation mentors

By the time the second five years of the Network began, the internal MOS, SMM, and OMSI evaluation departments conducted all of the evaluation work, and the summative evaluation work continued to be overseen by the COV. The Evaluation workgroup and Network Leadership met annually to decide upon the summative studies, allowing adjustments to be made based on changing Network needs and directions. Throughout this process, we needed to keep in mind that the studies should serve two purposes: (1) informing NISE Network decision-making, and (2) understanding the impacts of the Network for the knowledge of our funder, NSF, and the ISE field.

The first evaluation conducted under this structure, planned closely with Network Leadership, studied how partners used and felt about the Network’s main communication mechanisms. This study provided information to Network Leadership that helped them understand what partners wanted and make changes to better support participants. It also provided information about the communication structures of the NISE Network that could inform other ISE projects. However, it did not focus as much on the impacts of the NISE Network on individual participants. Therefore, in Year 7, the Evaluation workgroup, in consultation with Network Leadership and the COV, decided to divide evaluation work based on the NISE Network logic model (described in the Introduction). Using the logic model as a guide, the summative evaluation was divided into sections: SMM would focus on the impacts of the Networks’ biggest public deliverables (the NanoDays yearly programmatic event and the *Nano* mini-exhibition); MOS would conduct a three-year study on the impacts of NISE Network on participating professionals; and

OMSI would oversee a yearly partner survey and data mining process. For more detail about the actual set-up of the multi-institutional team, see the In-Depth Look at the end of this chapter. Embedded within this structure was the flexibility to make changes to what was studied each year, and the leader of the Evaluation workgroup still met with Network Leadership at least once a month to ensure that we understood any changes within the Network so our evaluations could accurately reflect what was happening at each point in time.

During Years 6–10, product development was still occurring, and NISE Network partners were also modifying existing products to fit their needs. This increase in the creation and modification of products meant there was also an increased need for formative evaluation, making it difficult for the Evaluation workgroup to provide formative data within a timeframe that could help workgroups and partners make product changes. However, over the course of the NISE Network, the Evaluation team had been working with partners to use and understand evaluation, thereby increasing their evaluation capacity. Because of this increased evaluation capacity, the Evaluation team and Network Leadership felt that workgroups and partners could take on formative evaluation on their own with support from the internal evaluators. Therefore, the Evaluation workgroup created Team-Based Inquiry (TBI), bringing together science inquiry and participatory evaluation techniques to help partners conduct formative evaluations on their own. TBI is described in more detail in Chapter 3 and in the TBI guide.⁵

Years 6–10

Further details about how the team was structured during this phase to deal with cohesiveness, objectivity, and flexibility can be found below along with information about how well this worked within the NISE Network:

MANAGEMENT CONSIDERATION	WHAT WE DID	WHAT WORKED	EVALUATION CHALLENGES
<i>Creating a cohesive evaluation across a multi-institution team</i>	<ul style="list-style-type: none"> • Internal evaluators led summative evaluation work. • The logic model was used as an aid to determine what evaluations were needed. • Integrating Team-Based Inquiry split the responsibilities for formative evaluation between internal evaluators and NISE Network workgroups. • Leaders of the SMM, MOS, and OMSI evaluation departments continued to meet at least once a month to discuss work. 	<ul style="list-style-type: none"> • Summative work was split among the internal evaluation institutions based upon the NISE Network logic model allowing a clear delineation of tasks. • With the addition of TBI, formative evaluators stayed embedded in their workgroups, acting as mentors for group members instead of evaluators. 	<ul style="list-style-type: none"> • Communication between evaluators on different summative studies was critical to ensure all aspects of the logic model were being covered. • Formative evaluators had to work closely with workgroups and each other to ensure consistent, adequate formative evaluation was being conducted through TBI.
<i>Generating an objective evaluation</i>	<ul style="list-style-type: none"> • The COV continued to provide oversight to the summative evaluation work. • The Evaluation workgroup continued to use multiple methods for summative studies. • Internal evaluators, with oversight from the COV, were embedded in workgroups to guide practitioners through the TBI process. 	<p>Embedded evaluators could ensure that workgroups were being fair and impartial in their TBI evaluations.</p>	<p>The level of rigor of the evaluations had to be relaxed so that practitioners could fit TBI into their existing work.</p>
<i>Remaining responsive and reflexive to the changing needs of the Network</i>	<ul style="list-style-type: none"> • Internal evaluators made decisions about summative work along with Network Leadership on a year-to-year basis. • Along with evaluators being embedded in Network Leadership groups, Network Leaders were embedded in evaluation sub-groups. • Regular memos and reports were produced to provide evaluation findings that could be used to inform Network decision-making. 	<p>Because evaluators and leaders were embedded in teams, evaluations could be designed and implemented as needs arose, and there was regular communication about studies.</p>	<p>Because of the need to plan ahead for budget and staffing, major changes to the course of the evaluation studies had to be carefully considered.</p>

What we learned through the changing NISE Network evaluation structure

A number of different team structures were used during the course of the 10 years of the NISE Network. While there were strengths and weaknesses to all of these structures, the final configuration worked best for the needs of the NISE Network. Yet in all of the cases, we feel that the success of the structure tended to reside in regular, ongoing communication among both Evaluation workgroup members and NISE Network workgroups through consistent participation on teams. These regular meetings allowed professionals to share expertise and learn from each other as well as ensured the usefulness of evaluations to the NISE Network. The inclusion of an outsider perspective through external evaluators and the COV also proved important as this was another way that evaluators could learn from each other and ensure that evaluations kept a view to the broader ISE field.



The most successful evaluations occurred when evaluators regularly attended team meetings but still retained some kind of outsider perspective. This was true of both the formative and summative evaluation work. Being embedded let evaluators hear what Network Leadership and workgroups were thinking and how their processes changed over time. Therefore, evaluators were able to suggest appropriate evaluation studies, determine when this work

was needed, and produce findings that were useful to the Network. When evaluators did not regularly attend meetings, they sometimes worked with old theories of action that led the evaluations to be not entirely representative of the work. It was also important for formative and summative evaluations to keep an eye to the needs of stakeholders who were not directly involved in the NISE Network project. External evaluators and the COV were able, at times, to provide this viewpoint and ensure that evaluation findings that were produced would be useful not only to NISE Network decision-making but also ISE and NSF decision-making. This continued exposure to multiple professionals allowed everyone to learn from each other and increase their capacity to do their work and conduct evaluation.

Another thing that we did, which worked for us and is probably unusual among informal science education projects, was embed members of NISE Network Leadership into the different evaluation subgroups during the last three years of the project. Network Leadership joined the evaluation subgroups so that they could have a better grasp of the evaluation studies. For evaluators, this meant that Network Leadership was able to provide further information about the way the Network worked. Finally, being embedded on teams allowed leadership to act as ambassadors for the evaluation throughout the Network as well as to NSF.

Across our work, we found that constant communication with internal stakeholders was key. In addition to the communication that occurred through the embedded evaluator in each team, we also practiced the following forms of communication:

- We created and shared reports, memos, and presentations with Network partners on at least an annual basis to help inform their work;

- We set up phone calls, in-person meetings, and other kinds of check-ins more than once a year to make sure that everyone understood what the Evaluation workgroup was doing and felt comfortable and confident in our work; and
- Even within our own subgroups, we relied on a range of methods for staying informed across institutions. For example, we held a variety of meetings between the various arms of formative and summative evaluations (weekly cross-department leadership calls, bimonthly Evaluation workgroup lead calls, monthly all-NISE Network evaluator calls, and biannual in-person meetings).

This level of communication took a lot of time and effort, but it allowed the Evaluation workgroup to stay informed about the changing evaluation needs of our partners. It also allowed NISE Network partners to stay informed about new findings.

What this means for other projects:

When planning an evaluation that uses evaluators from multiple institutions, consider:

- Using project documents related to impacts and outcomes, such as logic models and theories of action, as one way to determine how work is split among institutions.
- Having regular meetings, both virtually and in person, to make sure the different institutions have the same understanding of what is going on within the project.
- Placing at least one evaluator from each institution on each study to ensure that all institutions can provide input on and describe all the different evaluation work.

To create an objective evaluation, think about:

- Balancing the needs of internal and external stakeholders by providing data that informs project decision-making and the broader ISE field.
- Offering internal and external viewpoints by having evaluators or advisors who are more and less embedded in day-to-day project work.
- Using multiple methods for collecting data to provide the most holistic picture of the project's impacts.

When trying to manage a responsive and reflexive evaluation, consider:

- Planning evaluation studies throughout the course of the project instead of just at the beginning or the end of a project to allow them to be based on changing project needs and directions.
- Communicating findings with project stakeholders, including leadership, partners, and funders, in a regular and ongoing way to allow evaluation findings to inform project decision-making.



AN IN-DEPTH LOOK:

Figuring Out a Structure for Our Multi-institutional Team

When MOS, OMSI, and SMM joined together to work as internal summative evaluators for the NISE Network, one of the biggest struggles that we faced was how to split up the work while still conducting a cohesive evaluation. We tried a number of different ways of setting up our multi-institution team to complete this work, and we found that there were benefits and drawbacks to each option that we tried.

When we first took over some of the summative evaluation work in Year 4 of the Network, it was decided that each institution would have their own public impacts study to create and manage. In this model, each institution used only its own institutional staff to collect, analyze, and report on the data that came from their study. However, OMSI, MOS, and SMM shared information about their work during regular inter-institutional Evaluation workgroup meetings. A benefit of this model was that each institution had a clear charge and scope of work. In addition, every institution felt buy-in and ownership over their study. Issues that arose from this model included the fact that each institution did not necessarily understand

what was going on with the other studies because they were not involved on other institutions' Evaluation workgroups. This lack of understanding made it difficult for the evaluators working with different NISE Network workgroups to accurately and fully describe all of the evaluation work.

The Evaluation workgroup tried a second structure for their Year 6 and 7 work. In this structure, MOS, SMM, and OMSI worked together on a single study. Team members came from all the institutions, and they all had a hand in planning, data collection, data analysis, and reporting. A benefit of this structure was that workgroup members from all the institutions had a complete understanding of the evaluation

that was being conducted. Additionally, this structure allowed the team to leverage the expertise of workgroup members from different institutions. However, there were no clear guidelines for how the leadership of the work should be divided, causing confusion about how each institution should participate.

In Years 7–10, we struck upon a third model for distributing work. This model was a hybrid of the previous two. In this model, each institution had a subgroup that they were leading, and the primary support staff for the work came from the lead institution. Evaluators from the other institutions were also involved on each of the subgroups as *thinking partners*.

Thinking partners attended evaluation study meetings and provided feedback and advice on aspects of the study. A benefit of this structure was that each institution had ownership over an individual piece of work, making their scope and deliverables clear. Additionally, we could still ensure that workgroup members from all of the institutions fully understood all of the studies through the inclusion of the thinking partners. It was not always clear, however, how to best utilize the thinking partners on each evaluation project, and the lack of investment across projects by different institutions was still a possible issue.

In the end, this third model appeared to be the best compromise for our situation. We were able to involve all the institutions in some of the planning and implementation of all of the studies, and we were able to remain informed about everyone's work. Additionally, each institution had a clear charge and deliverables. However, there are situations where the other models may have been better. For example, giving each institution their own study without team members from the other institutions may have worked best if each of us had

expertise for conducting a specific kind of evaluation that matched the needs of the Network. Or, it may have been better for us to all have one single Evaluation workgroup if the scope of the Network had been smaller or more focused.

Through a hybrid model we were able to involve all the institutions in some of the planning and implementation of all of the studies, and we were able to remain informed about everyone's work. Additionally, each institution had a clear charge and deliverables.

CHAPTER 2:

Measuring Network Impacts

Marcie Benne and Marjorie Bequette

The Challenge: How can we define and measure impacts of an emergent and complex network in ways that are manageable, valid, and meaningful to a broad array of stakeholders?

Chapter Summary: NISE Net Leadership often said that Network partners learned by working together, by focusing in Years 1 through 5 on building the Network and in Years 6 through 10 on using the Network to engage the public. Through the entirety of the project, the NISE Network Evaluation team followed suit—learning by working with stakeholders in an environment that was new and complex, in ways that prepared our team to capture impacts within a large and dispersed system. This chapter explores:

- How the Network and Evaluation workgroup prepared to measure impacts;
- What frameworks helped ground the measurement of impacts;
- What decisions were made toward focusing on the Network’s intended impacts; and
- How the addition of research activities complemented the Network’s evaluations of impacts.

What we did

In a network of this size and scope, the choice of what to measure was never straightforward. Throughout the course of the NISE Network, Evaluation workgroup members collaborated with Network partners and Network Leadership to align goals, intended impacts, deliverables, and theories of action. Workgroup members evaluated deliverables based on goals so that partners could improve their products and practices and also document impacts. While the authors of this chapter were not involved in all NISE Net discussions about alignment and measures, here we provide an overview of how we saw evaluators working alongside Network Leadership to develop overarching frameworks of intended impacts and to determine how to measure these impacts.

NISE Net Leadership and Evaluators consistently cooperated to build on each other's strengths and make decisions in order to gain useful information for stakeholders. Throughout this chapter, we share some particular decision points related to measuring Network impacts in the hopes of informing future evaluation on large multi-institution projects. To illustrate how the NISE Network and Evaluation workgroup created measures of project impacts, this chapter describes the project in two major phases:

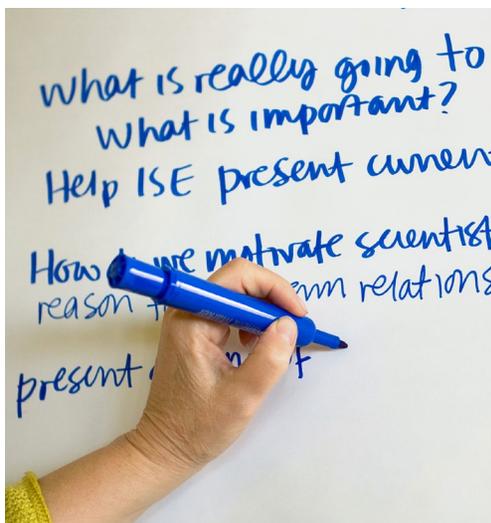
- Years 1–5: Network goals were new within an emerging and complex environment, which fueled diverse interpretations and experimentation with deliverables.
- Years 6–10: Network goals built upon strengths identified in Years 1–5, which allowed commonly shared interpretations to fuel the advancements of collective impacts.

Years 1–5: Learning to capture impact in an environment that was new and complex

Building relationships, experimenting, and aligning expectations

When the Network was new, the notion of achieving and measuring impact energized many Network partners, including exhibit and program developers. Between Years 1–4 of the NISE Network, Network Leadership and the Evaluation workgroup focused most of their efforts on experimenting with varied formats and content to understand what would work best for public participants of nano education activities and the partners who delivered those activities. The varying formats, content, and objectives needed for these audiences also required evaluators to experiment with a variety of outcome measures and methods. For instance, evaluators studied engagement, outcomes, and reach for demos, forums, exhibits, professional workshops, and partnerships using methods such as online surveys, written surveys, in-person interviews, telephone interviews, and naturalistic observations.





This experimentation started as an exciting period with partners developing educational activities while anticipating both intended and emergent impacts. However, for some of these professionals this became a challenging or even frustrating period as they realized the intricacies of evaluating

and interpreting the outcomes of the NISE Network project. Tensions were high when the Year 4 public impacts evaluation findings³ were reported and some of the interpretations presented by the Evaluation workgroup did not match the interpretations of Network Leadership. In particular, some Network stakeholders were taken aback when the Evaluation workgroup reported evidence indicating public participants were not learning about the risks of nanotechnology. The Evaluation workgroup found themselves in extended conversations with NISE Net Leadership, other workgroup members, and National Science Foundation (NSF) program officers discussing ways to approach goals, deliverables, measures, and interpretations of evidence in ways that were clearly connected, understood, and valued by stakeholders.

The Evaluation workgroup and Network Leadership embraced the tensions as learning opportunities and committed to developing a

clearer collective understanding about how nano could be shared with the public and how the Network could foster public learning by equipping professionals with training and activities.

Developing frameworks

As stakeholders continued to articulate shared intentions, Network Leadership decided to develop a set of frameworks to illustrate these intentions (see the Introduction for more detail). The Evaluation team and partners worked with Network Leadership to develop these frameworks and to understand their implications for evaluation.

- Network Leadership supported the development of a nano content map to illustrate key principles and interrelationships for learning nano across the Network (in Appendix A). The common nano education language within this map made it possible to discuss content priorities among impacts for professional partners and members of the public who participated in NISE Net activities and how to achieve the priority impacts.
- Because Network Leadership wanted to achieve collective impact through complex channels, a set of frameworks described the intended theory of action, activities, impacts, and reach for both professional and public audiences.
- Network Leadership also created professional development goals (Appendix B), and delivery and reach targets.

Together, these frameworks provided a roadmap for the development of future program and evaluation activities. Once these frameworks were created, it was easier for Network Leadership to communicate intended impacts more clearly and consistently to stakeholders.

POSSIBLE NETWORK IMPACTS

from perspective of centralized network leadership

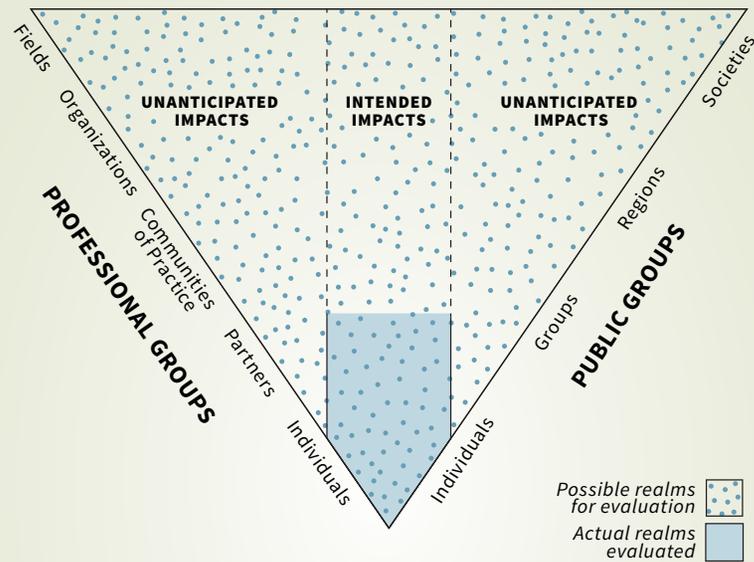


Figure 4. Possible network impacts

To evaluate the impacts of the Network, we thought hard about the scale at which to measure outcomes for public and professional audiences. For both, we could see the potential of studying impacts on individuals as well as broader groupings. Network leadership had explicitly or implicitly stated some intended impacts for all of the groups listed in Figure 1, but with less specificity as the group sizes broadened; thus, the impacts would be more diffuse and difficult to measure for the larger group sizes. Given the fact that evaluation resources for this project were not unlimited, and the notion that the Network needed to at least impact individuals in order to advance change among the broader groupings, we decided to primarily focus evaluation on measuring intended impacts at the individual level while still looking and listening for unanticipated impacts in the course of these studies. At the same time, the Network directed research resources to study questions related to some of the broader public and professional groupings.

Years 6–10: Measuring impacts

Even with frameworks, deciding what to measure in a network was not easy

The NISE Network used a model of adaptive leadership that aimed to be responsive to feedback from all directions and levels. While the common frameworks synthesized by the Leadership workgroup set a clear direction, it was understood that other impacts could emerge as the Network adapted to changing internal needs and external pressures. Because of this, the Evaluation workgroup spent considerable time discussing possible emergent impacts of the NISE Network, how they might be captured as a part of the evaluation, and whether they should be a focus of the evaluation.

The Network Leadership and Evaluation workgroups, including the COV, considered many possibilities of what could and should be evaluated to understand the Network impacts. We thought about studying emerging relationships and products. We considered evaluating the nature of the Network structure and operations as they shifted over time. We also considered evaluating how Network products prompted emergent changes within organizations or regions with deep nano involvement.

While these possible emergent impacts were intriguing to many stakeholders, conversations around evaluation priorities always returned to the importance of using resources to measure intended impacts over emergent impacts. The decision to measure intended impacts was possible, in part, because the centralized leadership had supported the documentation of the intended theory of action, target audiences, goals, impacts, and reach. This made it possible to plan evaluation studies in response to clear interest and direction from central stakeholders.

As a complement to the evaluation studies that focused on intended impacts, the Evaluation workgroup helped develop and institute Team-Based Inquiry, described in more detail in Chapter 3 and in the TBI Guide,⁵ as a flexible process for stakeholders at all levels of the Network to conduct formative evaluation on their own questions about their work, when they wanted to, and in a way that made sense to them.



Choosing units of analysis in a network was not easy either

Along with conversations about whether to measure emergent or intended impacts were conversations around possible units of analysis. In such a large network, it would have been possible to choose any number of units of analysis including, but not limited to:

- The Network
- Different kinds of involvement in the Network
- Geographic regions within the Network
- Partnerships within the Network
- Organizations within the Network
- Organizational sectors within the Network
- Professional sectors within the Network
- Professionals within an organization
- Public participants at different types of institutions

Once again, after many conversations around priorities, the stakeholders agreed that since the logic model identified individual professionals and individual public learners as the initial intended

target audiences, the evaluation activities should also center on the individual as the unit of analysis.

The decisions to study intended impacts and intended target audiences were just two of many decisions related to narrowing evaluation options. While the logic model framed the NISE Network impacts in terms of the public who participated in nano education activities (public impacts), and the partners

who created and delivered public products and/or participated in professional development opportunities (professional impacts), the challenge remained to think through how to evaluate these impacts with regard to factors such as specific groups (e.g., Spanish speakers and people with disabilities), site diversity, variations in activity delivery and engagement, depth versus breadth of data, sampling options, and the collective meaning of the data for Network stakeholders. Conversations around studying combinations of these and other factors were often sources of tension, innovation, and information that had to be considered, negotiated, and sorted by evaluation stakeholders.

Based on the set of NISE Network frameworks for Years 6–10, the Evaluation workgroup sketched a coarse overall evaluation plan with some flexibility for additions and adjustments within each of those years (see more information in Chapter 1). The sections below describe the general evaluation plan and study structure that we followed, with its focus on professional and public impacts to ensure that the NISE Network was adequately evaluated.

How the Evaluation workgroup measured the impact of the professional experience

The Years 6–10 Professional Impacts Evaluation subgroup was run by one institution, with support from Evaluation workgroup members from the other institutions and advice from Evaluation Leadership throughout, as described in the previous chapter. As noted above, the workgroup chose to study individual professionals because the NISE Network engages with individuals within organizations. One complication that remained after this decision was made was the definition of “professional,” which led to the question of how to include “scientists” and university staff members in this group when many products and practices were focused on ISE professionals. After much discussion among stakeholders, the Evaluation workgroup decided to conduct a multi-year study of all



individual professionals (ISE, academic, and science practitioners) with any Network participation. Since previous evaluations had shown that, in general, individual professional events and workshops were achieving their goals and in line with the logic model, Network stakeholders decided to expand the lens on professional participation to evaluate the broader treatment of “the Network” rather than focusing on individual events. Based on the NISE Network’s goals for professionals (see Appendix B), this multi-year study explored how involvement with NISE Net impacted an individual professional’s sense of community, learning about nano, and use of nano educational products and practices. There is much more to read about this study and findings on the NISE Network website (http://www.nisenet.org/About_Evaluation_Research).

How we measured the impact of the public experiences

The Years 6–10 public impacts evaluation subgroup was also run by one institution, with support from Evaluation workgroup members from the other institutions and advice from Evaluation Leadership throughout, as described in the previous chapter. Building upon evaluations conducted in Years 1–5, this subgroup planned three studies, which used common measures of learning whenever possible, based on the nano content map and previously used measures (more information in Appendix A). The subgroup collected data at multiple sites based on the particular questions of the studies and the need for broad sampling. A challenge for this subgroup was how to measure the impact on individuals when each institution was offering a different array of products. The description below of what was actually studied shows how the subgroup solved this challenge.

The subgroup decided to focus the first two public impact studies on the products that the participating public was most likely

to encounter—the small footprint exhibition, *Nano*, and the NanoDays events. These studies focused on measures of reach and learning. The reach side focused on understanding the number of people reached by these activities over the course of the project (extrapolations derived with data from all sites), while the learning side focused on the nano content individuals likely learned (outcomes derived with data from a small, carefully chosen number of diverse sites). The subgroup decided to focus the third study on the diversity of NISE Network public implementations, which included gathering and describing information about how institutions offered different NISE Network products, how many public participants were likely to encounter these multiple products, and how their learning might differ from participants who engage with just one product. More about all these studies and our findings can be found on the NISE Network website (*Nano* and NanoDays Summative Evaluation Reports: http://www.nisenet.org/About_Evaluation_Research).

Creating research studies

While the evaluation activities evolved to focus on individual intended impacts, NISE Network resources were robust enough to afford the creation of a Research workgroup to explore broader and emergent trends within the Network. Therefore, in Years 6–10, Network Leadership decided to fund several research studies to generate new field-wide knowledge about the ways in which science museum professionals and institutions can integrate new and developing science into the informal science education (ISE) field in the future. These projects—two run by the Museum of Science and the Science Museum of Minnesota through cross-institutional subgroups, and two run by outside researchers from the University of Wisconsin-Madison and SRI International—allowed for different kinds of questions to be

asked and answered than could be done through the evaluation of intended impacts. The foci of the four studies were the following:

- What does organizational change prompted by the NISE Network look like and what are the processes and conditions that facilitate or hinder this change?
- How do visitors use, interact with, and talk about the exhibit components within the *Nano* exhibition to learn about the relevance of nano to their lives?
- What is the nature of partnerships between university scientists and museum professionals in the NISE Network?
- What evidence does social media show of public engagement with nano, both inspired by the NISE Network and otherwise?



The Research and Evaluation workgroups had many joint discussions throughout Years 6–10 about possible intersections and differences among their findings. These research studies allowed the Network to contribute to theory in the ISE field and allowed the evaluation studies to remain focused on the extent to which intended impacts were achieved.

Ultimately, our collective decisions toward the approaches used to focus the evaluation studies on intended NISE Network impacts provided effective evidence of the extent to which impacts were

achieved. At the same time, we recognized our decisions had trade-offs such as not meeting requests from some stakeholders for additional evidence on public reach, broader measures of public impacts, links between short-term indicators of impacts and longer-term desired outcomes, and measures of impacts on the Network from directions other than the centralized Leadership workgroup. Readers who want to measure network impacts should also review how other networks approached these decisions and trade-offs.

What this means for other projects:

Before making final evaluation decisions for your project, try:

- Experimenting with possible goals, deliverables, samples, methods, outcomes, and indicators as a way to pilot ideas and see what will make the most sense for your evaluation work and the project.

When developing shared language about anticipated impacts, consider:

- How all stakeholders will benefit from contributing to the process and decision-making.
- Building on prior evaluation and research, including work in your area of focus, and work on educational networks and complex systems.

When evaluators are discussing what will be measured and not measured, try:

- Inviting stakeholders to share their perspectives and participate in the conversations.

When a project changes over the course of time, consider:

- Using techniques that will help stakeholders be continually aware of what is important for measuring impacts.

When planning a summative evaluation, remember that:

- A logic model and guiding document can help focus the study and provide a shared understanding, but network leaders and evaluators should actively consider and discuss the possibilities for studying unintended outcomes and broadening thinking to include different units of analysis as well.



AN IN-DEPTH LOOK:

Using TBI to Understand Stakeholder Conceptions of a “Worthy” Evaluation Study

Throughout the project, the Evaluation team worked to align expectations about the kinds of evaluation studies that would and would not be conducted. One way we approached this was to conduct a Team-Based Inquiry (TBI) study in Year 7 to understand what four of our stakeholder groups (the Committee of Visitors, other workgroup leaders, Network Leadership, and the Evaluation workgroup) felt constituted a “worthy” evaluation study (“worthy” meant important to do given the finite resources of the Network).

TBI, described in more detail in Chapter 3, is a process developed by the NISE Network that supports professional inquiry to inform practitioners’ decisions about their work. The Evaluation workgroup’s TBI revealed that different stakeholder groups had different ways of articulating what makes a study worthy. The results were as follows:

- Network Leadership members tended to mention broader aspects of the NISE Network related to the theory of action.
- Workgroup leaders tended to focus on particular NISE Network activities they wanted to understand better.
- COV members focused on methodological characteristics.

Based on these findings and significant additional dialogue, the Evaluation workgroup organized their methods and structures to support the production of worthy studies. For instance, to address the broader aspects of the NISE Network, the Evaluation workgroup shifted the focus of the subgroups slightly to the final team structure: one subgroup studied professional impacts, one studied public impacts, and one documented the Network Leadership’s theory of action and primary initiatives over time. To address workgroup leaders’ interest in better understanding their activities, the Evaluation workgroup further refined and promoted the team-based inquiry method so leaders could get the data they wanted

when they wanted it and use it to inform improvements to their work.

In terms of methods to support worthy studies, the team also appointed an evaluation subgroup to focus on a robust, consistent annual survey that would support both workgroup and Network Leadership needs. This annual survey was available for workgroups, evaluators, researchers, and others to ask their own questions of Network partners. The responses to those questions were returned as quickly as possible for use in decision-making about their NISE Network activities. Additionally, the Evaluation workgroup built in structures with redundancies and depths in staff to ensure succession planning and knowledge breadth, and

the Evaluation workgroup hired a team member with specific statistical expertise to support studies as needed.

By making sure to gather information from our main stakeholders, the Evaluation workgroup was able to craft evaluation plans that included worthy studies and would work for as many of our partners as possible. This work indicates the importance of gathering information from your own stakeholders about what they consider a worthy evaluation as it could help project evaluators to determine how to prioritize and plan for their studies.

The Evaluation workgroup's TBI revealed that different stakeholder groups had different ways of articulating what makes a study worthy. Based on these findings and significant additional dialogue, we organized our methods and structures to support the production of worthy studies.

CHAPTER 3:

Adapting Methods for Network Use

Sarah Cohn, Juli Goss, Liz Rosino Wright, and Gina Svarovsky

The Challenge: How can we effectively measure the impact of a network across a range of individuals, organizations, and types of educational products without overburdening partners through their participation?

Chapter Summary: Over the years, we used various data collection methods to measure the public and professional impacts of the NISE Network, such as interviews, surveys, focus groups, observations, and mining existing data. There were also a few methods, in particular, that we developed and/or adapted to best support the study of the Network and its complexities, given the national scale of the Network across a variety of geographic locations, program formats, and stakeholders. This chapter reviews:

- How we distributed data collection for one-day events;
- How we estimated reach beyond ticket sales;
- How we used a network-wide survey; and
- How we employed Team-Based Inquiry.

What we did

The size, scale, and variety of activities of the NISE Network provided many challenges to collecting evaluation data. In many cases, it pushed us to adapt basic methods for this national scale, all while being mindful of not overburdening our partners and stakeholders. For example, there was a need to collect data at one-day events happening simultaneously around the country. We knew that in order to understand the full impact of these events, one data collector stationed at the entrance of just one event would not be enough. There was also a need to measure the audience reach of partner activities happening throughout the year at museums and universities. We could not assume that everyone visiting the organization was also encountering nano because that would over-inflate the number of people reached. Other challenges included not overburdening professional partners with too many requests



to participate in studies while also trying to support the growing formative evaluation needs of the Network. The following sections describe these selected challenges in more detail, how we handled them, and how the method we used could be helpful to others.

Distributing data collection for one-day events

THE CONTEXT:

Since 2008, NanoDays events have occurred simultaneously within at least 100 different organizations during a week-long period each spring. Every organization's event was different. Some of the locations had one area devoted to NanoDays activities, while other locations had activities scattered throughout their institutions. For the summative evaluation of NanoDays, we knew we could not be everywhere in one particular institution at once, let alone at every NanoDays event across the nation, so we needed to develop methods to collect data within and across diverse contexts.

WHAT WE DID:

For the summative evaluation of NanoDays, we collected data at nine events held within both small and large organizations. To ensure that many of the logistics surrounding this event had been worked out by an institution, we knew we wanted to select organizations that had hosted NanoDays before for our sample. In addition, we made sure to include within our sample organizations of different sizes and types, which would be more representative of the Network overall. We worked closely with a main contact from each partner organization to learn the layout and scope of their NanoDays event to make sure we were prepared for data collection. In order to collect data across the event as a whole and in-depth with a subset of individuals, we developed a data collection plan that incorporated surveying as

many people as possible while also allowing for interviews with some participants. Ultimately, we ended up sending out teams of three evaluators to each site with each team member focused on a different aspect of data collection:

- One evaluator concentrated on collecting surveys from as many adults as possible;
- A second evaluator collected both surveys and brief interviews from adults; and
- A third evaluator conducted interviews with children.

By collecting data as a team, we were able to capture responses that one method of data collection might have missed. Unlike many ongoing informal science learning experiences, the nature of our one-day events meant that we only had one opportunity to collect data, so using this system of collecting data was key.

WHAT THIS MEANS FOR YOU:

Whether it's NanoDays or another special event, collecting data through multiple methods and in multiple locations will help provide a better understanding of the impact. Finding an optimal balance between the breadth and depth of your data can help you create a plan for deploying the data collection resources that are available to you. For further details about our data collection, see our report on the Summative Evaluation of NanoDays (http://www.nisenet.org/About_Evaluation_Research).



Estimating reach beyond overall visitor attendance data

THE CONTEXT:

NISE Network partners across the country used a range of educational products in a variety of formats. In addition to providing free educational materials online, the NISE Network disseminated programmatic kits of materials to hundreds of museums and universities each year since 2008. In 2011, the Network deployed the first of 93 copies of the *Nano* exhibition. Because each partner organization was encouraged

to use any combination of these materials with their audiences, we knew it would be difficult to determine how many individuals were being reached by NISE Network products. However, our team needed to understand NISE Network's public reach not only so we could report it to NSF, but so that Network Leadership and product developers could make decisions about how much time and what resources to dedicate to these efforts. With hundreds of organizations using NISE Network materials in different ways, we wanted to be more accurate than relying on gate revenue or ticket sales. To do this, we needed a systematic way of counting when each implementation of NISE Net educational products with visitors at the various partner institutions looked different.

WHAT WE DID:

We knew we could not count every single person at all NanoDays events or exhibition host sites across the country, so we had to estimate. The bad news was that an estimation would not be 100%

What counts as counting?

When we asked partners to count the number of visitors interacting with individual programs, we provided these guidelines as a part of a virtual training:

- For programs and exhibits where visitors are likely to wander in and out (e.g., cart demonstration, exhibit, poster display), use the clicker to count every person who participates in the activity. Someone is counted as a “participant” if they touch something or pay attention for five seconds or more.
- For programs in spaces with a fixed capacity (e.g., stage presentation, theater), estimate the number of participants based on the capacity of the space. Someone is counted as a “participant” if they watch the presentation for five minutes or more.

We had also pilot tested using an incentive method. For the incentive method, we provided the following instructions:

- The educator facilitating the cart demonstration should give a small incentive to each person visiting the cart such as the temporary tattoos provided. In order to arrive at a count, subtract the number of incentives you have at the end from the number you started with to determine the number of people who interacted with the cart.

During pilot testing, we found that this method was not accurate, as most educators were not able to give the incentive to all participants either because they forgot or because they were engaged with another visitor, so we decided not to use the incentive method for our study.

accurate, but the good news was that having a systematic process resulted in a more accurate estimate than merely reporting the annual attendance of every institution using NISE Network materials. Therefore, our methods aimed to estimate the number of people

actually interacting with the NISE Network products rather than only being in the same building. What follows is a description of how we did this for two of the NISE Network’s major initiatives.

NanoDays events

Because NanoDays events are different at each site, we began figuring out how many people could be reached by the components of the event—the individual programs. To do this, we asked institutions of varying sizes and types to count the number of people interacting with an array of randomly selected programs. Small museums, large museums, and universities participated in this part of the study, and we provided training on ways to count the number of people at different activity types. After we knew how many visitors interacted with a program type at a specific institution size, we asked all institutions to report which programs they included in their NanoDays event. Then we extrapolated to determine the number of times that NISE Network products reached a member of the public. Because NanoDays events often have multiple educational offerings presented at once and an individual will likely encounter more than one program, this method for estimating reach of events had a limitation in that we were only able to report the number of “encounters” or a duplicated count. We were not able to determine the number of individual people reached, but rather the number of times there was an interaction between NISE Network materials and an individual.

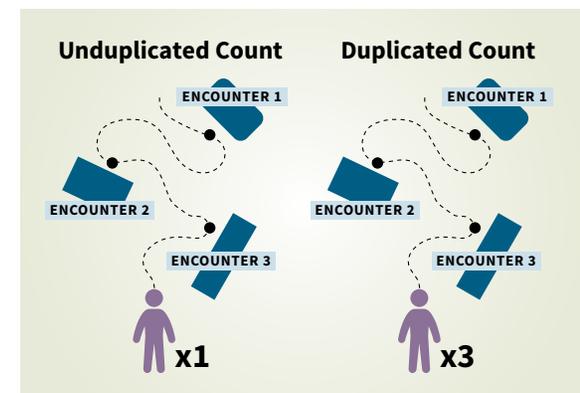


Figure 5. Unduplicated vs. duplicated count

Nano mini-exhibition

We determined the estimated reach of the *Nano* mini-exhibition as a part of the summative evaluation for the exhibition. This summative study asked professionals at the seven museums that initially hosted the exhibition to observe and record the number of visitors coming in contact with the exhibition at half-hour periods during morning and afternoon times for multiple weekdays and weekend days. This gave us the total number of people in contact with *Nano* per half hour. Then we asked each museum for their daily attendance, which allowed us to calculate how many visitors came to the museum per half hour. By comparing those rates, we estimated the average percentage of visitors who saw *Nano* on a weekday and a weekend day.

For example:

Average # of visitors who see <i>Nano</i> per weekday half hour	26 people
Average estimated weekday attendance per half hour	30.52 people
Average percentage of visitors who see <i>Nano</i> on weekdays at one site	85.20%

We averaged the weekday and weekend percentages to determine the average yearly projection of visitors interacting with *Nano*. We did these calculations for each of the seven initial sites, and then applied the most conservative yearly projection to all of the other sites hosting the *Nano* mini-exhibition. One limitation of this method for estimating reach was that we wished the Evaluation workgroup could have collected the number of people per half

hour at more times throughout the year to capture any notable variability. However, we felt that this method was more accurate than simply reporting the total number of visitors to all *Nano* host sites, as it relied on the percentage of visitor attendance based on collected data. This estimation was reported as an unduplicated count or the number of individual people who visited the mini-exhibition.

WHAT THIS MEANS FOR YOU:

You could develop a counting estimation for individual exhibits or other activity types. This would be useful for organizations or evaluators who are a part of multisite projects interested in methods more accurate than a self-report or ticket sales. For further details on this data collection method or calculation procedure, see our memo about all of our reach estimation work (http://www.nisenet.org/About_Evaluation_Research).

Using a network-wide survey

THE CONTEXT:

As part of ongoing evaluation projects, research studies, and product or community development, there were many questions that the Network wanted to ask partners. There was also a need for the Network to collect administrative-type information from partners for grant reporting. Because of this, partners were being asked by multiple people, multiple times per year to provide information and fill out questionnaires. After a few years, we started to hear from partners about their fatigue related to surveys and other requests for information, and we worried about this getting worse over time. To reduce the burden on partners, there was a need to consolidate many of the questions into one annual data collection method.

WHAT WE DID:

A survey team was created to coordinate the collaborative development and implementation of a single annual online survey. About three months prior to launching the survey each year, the relevant stakeholders (evaluators, researchers, developers, leadership, etc.) were asked to submit a description of the type of information they were interested in learning from partners.

The team compiled and refined the survey through an iterative process to focus on thematic flow, consistency of question styles, and length. For example, if three stakeholders each wanted to ask a few slightly different questions about museum and scientist partnerships, the survey team worked to bring these stakeholders together and suggest ways to merge the questions in a succinct and logical way that worked for everyone.

Once the survey was finalized, it was left open for about three weeks each fall. During this time the team focused on recruitment and delivered several follow-up reminders with partners to increase the response rate. Once the survey closed, the team cleaned the data and compiled and shared a summary report that included information about responder characteristics and frequency tables for each of the close-ended survey questions. Interested stakeholders could then request the actual data to any or all of the survey questions for their own analysis. Some of the data were used by the evaluation and research teams who incorporated it into their studies and other data were used by Network Leadership and



workgroups who analyzed it as part of their Team-Based Inquiry (TBI) studies.

WHAT THIS MEANS FOR YOU:

The collaborative development model for a survey could be used when there are many interested parties wanting slightly different types of information from the same group of people. For example, a survey sent to an organization's members could include questions to serve different departments'

needs. In this way it benefits both the organization in having an efficient and widely useful tool and the respondent in not having to fill out multiple surveys or other forms of feedback.

Employing Team-Based Inquiry (TBI)**THE CONTEXT:**

As Years 1–5 came to a close and the next grant for Years 6–10 began, it became clear that NISE Network partners had more questions and products that needed formative evaluation than the Evaluation workgroup could manage. Moreover, with encouragement from NISE Network, Network partners and participating institutions were adapting, modifying, and developing their own nano educational products. The Evaluation workgroup needed a way to support the wide range of evaluation efforts taking place across product development, community engagement, and the broader Network.

WHAT WE DID:

We drew on various theoretical perspectives, including action research,⁶ practitioner inquiry,⁷ learning-focused and participatory evaluation,⁸ and evaluation capacity building⁹ and grounded it in the inherent nature of science museums—scientific inquiry—to develop, implement, and train practitioners on a familiar cycle of question–investigate–reflect–improve. This process became known as Team-Based Inquiry (TBI), which is “a practical approach to empowering education professionals to get the data they need, when they need it, to improve their projects and practices and, ultimately, more effectively engage public and professional audiences.”⁵ For examples of what this looks like in practice, check out the TBI Guide and other resources in the box in this section.



Beginning in Year 6, those individuals who were directly funded to engage the Network community and develop educational products utilized TBI to conduct their own formative evaluations. In Years 8, 9, and 10, the Network focused on sharing

this model beyond those who received direct grant funding and offered professional development in a variety of ways. We presented and shared this information with professionals from museums and universities through conference sessions and workshops, intensive cohort trainings, videos, online materials, and a written guide that was available online or in the NanoDays kits.

The TBI development and implementation team included evaluators, program developers, and facilitators. Because TBI is intended to be a practical approach to help practitioners use data to make decisions, we felt it was important that the team include a wide range of professionals. Additionally, everything that was developed and disseminated to the broader Network community was reviewed and approved by the Evaluation workgroup and Network Leadership.

WHAT THIS MEANS FOR YOU:

TBI helped maintain the quality of educational products and practices developed by the Network, and allowed the Evaluation workgroup to focus on other efforts measuring the impact of the broader Network. As we have disseminated the model of TBI, many professionals have successfully conducted small-scale studies to inform their practice. Future network evaluators of large-scale, multisite projects could consider including TBI as one approach for the formative evaluation phase, while freeing up some resources to focus on evaluating other network efforts.

For further details, see our Team-Based Inquiry Guide and related materials and videos online:

http://nisenet.org/catalog/tools_guides/team-based_inquiry_guide

<http://nisenet.org/catalog/team-based-inquiry-training-videos>

What this means for other projects:

When collecting data for a one-day event, think about:

- Deploying multiple data collectors as a team in order to capture feedback across the event.

When reporting the number of people reached, consider:

- Using either a duplicated or unduplicated count depending on what is the most appropriate method for your project. If you have the same type of program across multiple sites, consider extrapolating based on a systematic estimation from a subset of locations, rather than a self-report of every location.

When developing a survey collaboratively, try:

- Combining questions that a respondent would consider redundant, and including extra time during survey development so that you will be able to negotiate between teams.

When supporting educators in conducting formative evaluation, propose:

- Using a Team-Based Inquiry approach, drawing on the resources mentioned above.



AN IN-DEPTH LOOK:

Evaluating with Diverse Audiences

The individuals and organizations that made up the NISE Network Evaluation workgroup represented a range of expertise. While we worked to evaluate the impacts of the Network, we shared our collective knowledge on everything from rating scales to focus group protocols to analysis software. Conducting evaluation with diverse audiences is another example of where we have learned from one another not only about specific methods, but also about an overall approach to evaluation.

Below we share some of what we learned by working together to include diverse audiences on our evaluations and several of the modifications we made to existing protocols as a part of this work, in hopes that other evaluators will consider ways to be inclusive of a range of individuals in future studies.

Since its inception, the NISE Network has been committed to developing and sharing nano educational products with a wide range of public audiences. While there was a specific work group committed to efforts related to diversity and inclusion, all educational product and community developers worked to achieve these aims. As evaluators, we felt it was

important to reflect this context and be as inclusive as possible in our evaluations as well.

One of our studies that illustrates this commitment to being inclusive is the summative evaluation of the *Nano* exhibition. As detailed in the section of the evaluation report titled “Nano Mini-Exhibition Audiences,”¹⁰ the development team sought to optimize use of the exhibition for all audiences and included elements to specifically address visitors with a range of abilities and those who speak Spanish. For example, the exhibition was guided by principles of universal design,¹¹ which includes adding multisensory access points. Labels were

also designed using side-by-side English and Spanish. In addition, expert advisors who were a part of these audiences were involved in the formative evaluation and review process. To evaluate the results of these efforts, the Evaluation subgroup conducting the study was purposefully assembled to include someone who had professional experience conducting evaluation with people with disabilities and a bilingual/bicultural evaluator who had professional experience working with Hispanic audiences. Other considerations for the study included being mindful of the national scope of the Network and the potential regional differences to ensure appropriate data collection instruments

and to allow time for additional pre-communication with partner sites to plan the data collection.

While further details about how we conducted this study and our findings are provided in the *Nano* summative evaluation report,¹⁰ we would like to highlight the following aspects of the data collection, analysis, and reporting process.

During the exploratory study of Hispanic audiences

- Bilingual/bicultural evaluators translated all survey instruments, collected all data, and conducted analysis of any data collected in Spanish.
- During the observation portion, data collectors noted the language used by the group in the exhibition and asked during the interview which language the participants preferred to use for the interview.
- We made sure to aggregate data within an institution and not across US locations in order to recognize and represent regional diversity accurately.
- Finally, when reporting these findings, we reinforced that this study was not designed as a comparison between Hispanic and

non-Hispanic audiences, and should not be interpreted as such.

During the exploratory study of visitors with disabilities

- We used universal design to guide our evaluation in order to determine how the exhibition promoted or limited physical, cognitive, and social inclusion.
- We recruited visitors with a range of abilities in order to understand if the exhibition was successful across various types of disabilities. In particular, we made sure to include visitors who are blind or have low vision as the audio descriptions were of specific interest to the NISE Network development team.
- We asked follow-up questions about our observations to make sure that we were correctly interpreting evaluation participants' exhibit use.
- Finally, we made sure to convey to participants that this was a study of the

design and effectiveness of the exhibition rather than a critique or judgment of individual abilities.

While some members of our team had prior experience conducting evaluation with these audiences, we felt like this evaluation was positive because we learned

Since its inception, the NISE Network has been committed to developing and sharing nano educational products with a wide range of public audiences. As evaluators, we felt it was important to reflect this context and be as inclusive as possible in our evaluations as well.

from one another, we were responsive to the Network and the museum visitors, and we were able to discuss outcomes not always documented in evaluations that are conducted for our field. For more information on cultural competence in evaluation, see the statement produced by American Evaluation Association.¹²

CHAPTER 4:

Creating Ethical Evaluation Practices for a Large Network

Elizabeth Kunz Kollmann

The Challenge: How can we ensure that our evaluation participants and their data are treated ethically while also getting the Network the data it needs to make decisions about its work?

Chapter Summary: Our task as the Evaluation workgroup was to craft methods for ensuring that everyone involved in any NISE Network-related evaluation process was treated respectfully and that collected human subjects data were adequately protected. These concerns related to professional evaluators, practitioners who conducted formative evaluations through Team-Based Inquiry (TBI), and participants in our evaluation studies. This chapter illustrates:

- What we did to make sure that data were ethically collected;
- What protections we put in place for evaluation subjects; and
- What safeguards we established for data collectors who were also ISE professionals.

What we did

The Evaluation workgroup wanted to make sure that we provided data to stakeholders or aided them in collecting their own data so that improvements could be made to the NISE Network deliverables. However, we also wanted to make sure that we ethically treated our evaluation participants and their data in the process. Therefore, a big task was figuring out how to balance data needs while protecting our evaluation participants.

Part of protecting our evaluation participants and data collectors was crafting and managing approved human subjects protections and protocols overseen by an Institutional Review Board (IRB) that would govern the ethical collection and use of data. While this is not required of all informal science education (ISE) projects, it is a requirement of projects funded by the National Science Foundation (NSF) such as the NISE Network. In our case, the Museum of Science (MOS) has their own IRB, so we created protocols for the entire NISE Network Evaluation including front-end, formative, and summative evaluation work and had them approved through the MOS IRB. For us, crafting IRB protocols was helpful because they provided clear rules about:

- Gathering informed consent from evaluation subjects before data collection commenced;
- Ensuring that necessary data was collected from participants through methods that were as noninvasive to evaluation participants as possible; and
- Handling data in a way that protected participants during the data collection, entry, analysis, and reporting processes.

However, there were added difficulties in crafting a single set of NISE Network IRB protocols because of the size and scope of the NISE Network. These issues included the fact that:

- Data collectors came from multiple institutions;
- Non-evaluators were going to be involved in collecting some of their own data;
- Our direct partners were some of our main evaluation participants; and
- People or institutions might be asked to participate in multiple studies over the course of the project.

According to the US Food and Drug Administration:

“The purpose of IRB review is to assure, both in advance and by periodic review, that appropriate steps are taken to protect the rights and welfare of humans participating as subjects in the research. To accomplish this purpose, IRBs use a group process to review research protocols and related materials (e.g., informed consent documents and investigator brochures) to ensure protection of the rights and welfare of human subjects of research.”¹³

Even if an IRB is not required as a part of a project that you are working on, considering ethical treatment and protection for participants, as well as your data collectors, is an important part of conducting research and evaluations with human subjects. This is because you may ask research and evaluation participants to provide

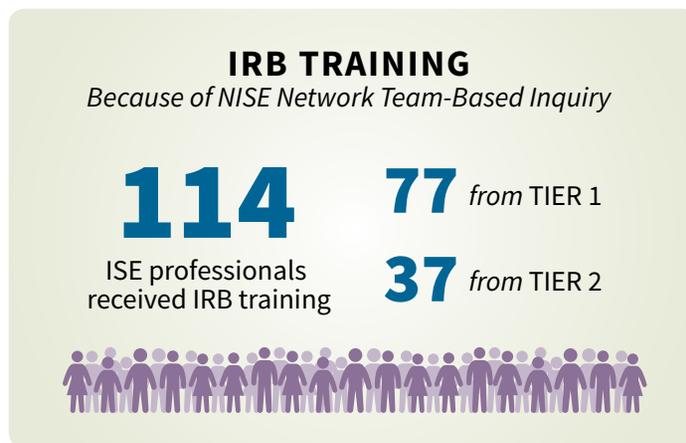
personal information about their viewpoints related to sensitive subjects such as health or security, and you should design your study to minimize the risk that your subjects will feel uncomfortable. Working with professionals as subjects and data collectors also leads to additional concerns that need to be considered. By asking professionals to provide or collect data, it is possible that you are putting their future careers or ability to receive resources in jeopardy. With both public and professional participants, you are asking them to give up some of their time in order to provide you with information, so you need to think about how to balance evaluation with participant needs. Below we describe how our team dealt with these concerns by managing the:

- Training of evaluators and non-evaluators;
- Collection of data from and with colleagues; and
- Collection of data from the same sites and people over time.

Training evaluators and non-evaluators about human subjects protections

It was important for us that all individuals who collected or handled data treated our evaluation participants with respect and followed ethical guidelines as outlined by groups such as the American Evaluation Association.¹⁴ This was made more complicated within the NISE Network because once Team-Based Inquiry (TBI) was put in place around Year 7, non-evaluators were now the ones who were

conducting much of the formative evaluation of their deliverables (see Chapters 1 and 3). Therefore, it was important for us to think about the kinds of materials and protocols that we could put in place to ensure data collectors and handlers followed consistent, ethical practices. This process was especially challenging because we were working with so many individuals for whom 1) evaluation was not a normal part of their jobs, and 2) we could not take them away from their regular work for too long. Thus, we had to come up with a number of different ways to train and provide information to all of our data collectors to ensure that they were consistently following the approved IRB procedures.



One thing that was required by our IRB, but that may not be required for all projects similar to ours, was that everyone who was collecting or handling data needed to go through full human subjects training. Human subjects training is seen as necessary for many projects that involve collecting data from people as a way to ensure that those involved in this work have a common understanding about risks to human subjects and how to avoid them. An added benefit of going through this training is increased capacity of project members to advocate for the protection of research and evaluation participants. For the NISE Network, getting full human subjects training involved participating in an online course. There are multiple versions of this training, but the version that the NISE Network used was from the National Institutes of Health Office of Extramural Research.¹⁵ This human subjects training is free and introduces research ethics,

obligations of the investigator, and methods for protecting the rights and well-being of human subjects.

One thing that we struggled with was that this training can take up to two hours to complete. This duration can be a prohibitive amount of time for some projects and was also considered problematic for the NISE Network. We tried to come up with different ways for our data collectors and handlers to obtain the training they needed without completing this course. However, these efforts ended up not being acceptable to our IRB (see the In-Depth Look at the end of this chapter for more information). Therefore, we decided that a good way to alleviate at least some of the burden of the training was to limit who had to complete the course. In the case of the NISE Network, we decided that for our IRB, full human subjects training was required of anyone:

- Recruiting and obtaining consent from potential evaluation subjects;
- Collecting data from or about evaluation subjects; or
- Entering and analyzing data.

Together with our IRB, we decided that individuals did not need to complete this training as soon as they joined the NISE Network. Rather, anyone who participated in these activities could complete the online human subjects training within a year of joining the project and before taking part in any evaluation activities. This gave our partners the chance to fit the training in when it was most convenient to them.

Even though we made people undergo full human subjects training, we knew that the IRB document that laid out the overall NISE Network protocols for collecting and handling data was long and not

very digestible for either evaluators or non-evaluators. Therefore, we created a shorter data collection and handling guide¹⁶ that was specific to the NISE Network. This document was based on one created by the Oregon Museum of Science and Industry (OMSI) to help their own data collectors and included:

- Sample scripts and information about how to get consent from subjects when collecting either 1) anonymous/confidential data through interviews or surveys, or 2) identifiable data through videos or audio recordings;
- Information to help data collectors think about what data they actually needed to collect, along with encouragement to collect identifiable information such as names and email addresses only when absolutely necessary; and
- A description of how to de-identify data when entering and processing it.

We found that it was helpful to use both the online course and reference document when training data collectors and handlers. The full human subjects training provided a universal view of why it is important to follow human subjects protocols and why IRBs were established by the US government in the first place while the NISE Network data collection and handling guide put all of our protocols for the NISE Network project in one place.

Collecting data from and with colleagues

Another issue that we ran into was concerns about informal education professionals participating in our studies and being active in analyzing data. This group of professionals was important to us because we wanted to collect data from them about their



experiences with the NISE Network to better understand the impacts of the Network, but also because we wanted them to be involved in analyzing formative evaluation data through TBI efforts (see Chapter 1 and 3 for more information). However, concerns arose about professionals being evaluation subjects and acting as data collectors because we recognized that they might feel coerced to participate in evaluation activities or feel that they could not be completely honest due to potential impacts on their jobs or professional relationships. Therefore, we put some protocols in place to help these groups feel comfortable about and confident in the evaluation.

To ensure data collectors were fair to the fellow professionals that they were collecting data from, we made sure that groups collected identifiable information, such as names and institutions, only if this information was absolutely necessary. This allowed professionals' data to generally remain anonymous. However, one issue that we came across with the NISE Network was that some teams were collecting data for NSF reporting purposes that had to include some

identifiable information. For example, the NanoDays kit team had competing data needs because:

- They needed data about how each institution used their NanoDays kit so that they:
 - Could report this to the funder, and
 - Use these data to assist with a competitive allocation of future NanoDays kits and other network resources such as professional development opportunities.
- They wanted data for their TBI studies to help them think about how to design future kits.

In order to enhance informal education professionals' comfort responding to questions about how they felt about current or future kit activities, we devised ways to separate these data from the reporting questions so that responses would remain anonymous. We did this by creating two separate surveys: one survey with the required reporting information and a separate survey with the evaluation questions so that names and institutions were not connected to the evaluation data. We also added consent language letting participants know that the evaluation data would not be used to make decisions about whether they would receive future NISE Network products or professional development opportunities.

When asking informal education professionals to analyze data along with their peers through data reflection discussions during the TBI process (see Chapter 3), we were worried that some members of the project teams might not feel comfortable because the discussions were often about deliverables that other members of their team had created. Because we recognized that discussions may reveal negative

results, we gave professionals the opportunity to opt out of aspects of the analysis process. These protocols were written into our human subjects materials and verbalized to all of the NISE Network workgroup members. However, we found that no one requested to be excused from the TBI data analysis process. It is possible that this was because most team members had worked with each other for years prior to the introduction of TBI so they already felt comfortable with each other. It is also possible that people did not opt out of analysis because of NISE Network's intense interest in using data and feedback to improve products and practices.¹⁷ Therefore, as a whole, the project team members might have been more open to data-based criticism. However, it is also possible that professionals felt pressure to participate despite being told these discussions were voluntary. In any event, it is important to maintain open lines of communication with practitioners involved in data analysis to ensure that they feel comfortable with the process.

Collecting data from the same sites or people over time

The extended time scale and size of the NISE Network posed a problem for the Evaluation workgroup because we wanted to collect data from the same sites and/or individuals multiple times over the 10 years of the project. After a few years working on the NISE Network, we found that some individuals complained that the NISE Network was asking them to provide data multiple times per year, and that they didn't understand why there were so many different data requests. These complaints were in part because we were not only requesting that partners participate in evaluations, but the NISE Network was also asking partners who received NanoDays kits, mini-grants, or an exhibition to fill out additional

reports about how they used and what they thought of these resources. Due to the issue of evaluation fatigue, we were beginning to see a decrease in our evaluation response rates. Therefore, it became imperative that we come up with methods for reducing data collection burdens on our partners.

To solve this issue, one thing we did was to create an annual survey that included questions from multiple stakeholders at once (more information can be found in Chapter 3). When the survey was established in Year 8, all of the NISE Network workgroups were told that this was their only opportunity to ask questions of all NISE Network informal education professionals for the year. Some benefits of the annual survey were that it allowed us to:

- Better control the number of times partners would get feedback requests each year;
- Ensure that partners were not being asked the same questions multiple times; and
- Lead teams to existing data sources to answer their questions instead of having them collect additional data.

A drawback of the single survey was that we sometimes had to modify or leave questions off because of concerns about making the survey too long. Also, it was sometimes difficult for workgroups to know the questions they would need on the survey when the planning process started months in advance. Nonetheless, we determined that it was still more desirable to have a single partner survey each year.

Another way that the Evaluation workgroup tried to alleviate burdening partners with data requests was to create a list of the

partners who were involved in our different evaluation studies. This allowed us to keep track of who had been asked to take part in different studies, the kinds of data that they had provided, and when these data had been requested. Thus, whenever a new study was designed or a new group came to us with a request to do research on NISE Network partners, we could check the list to make sure that no one institution or individual was being asked to provide data too many times. Occasionally, this meant that we requested certain evaluators or researchers not to approach a potential subject. In other instances, this meant that we let potential subjects know we were aware they were involved in other studies and appreciated their ongoing involvement, but were approaching them for an additional study because we felt their perspective would be important here as well.

Tracking subjects between evaluation studies was not difficult within the Evaluation workgroup. However, it was more challenging to integrate studies from groups outside of evaluation such as those conducted by the Research Team, which was created in Year 6 to generate broader understandings of NISE Network products and practices for the informal science education field (see Chapter 2 for more details). Because members of the Research Team did not meet regularly with members of the Evaluation workgroup, we did not always know when and with whom data collection was going to occur, and they did not always know about the participant tracking system. To improve communication, it might have helped to have more regular meetings between our groups and to have a way of more formally introducing the participant tracking document to the Research Team so they would understand its purpose and importance.



Finally, as the Network matured and increasingly more data were collected, the Evaluation workgroup recognized that we might be able to answer some of our questions with existing data instead of re-asking partners. This included both data that we had collected as a part of our evaluation studies as well as data that were collected for other purposes, such as NSF reporting. Therefore, we set up the data mining subgroup to explore what information we already had that could be used to answer different evaluation questions (see the Introduction for more information). This allowed us to make sure that our partners weren't being asked to answer the same or similar questions over and over again unless it was necessary for understanding the impact of the Network over time. As described above, this also helped us point other teams to existing data sources that could answer questions about their work. For example, one group decided that instead of asking partners what they desired from the group, they would mine past emails for this information.

What this means for other projects:

When using non-evaluation professionals as data collectors and handlers, consider:

- Providing different kinds of trainings and reference materials to meet the needs of individuals with varying levels of expertise about evaluation.
- Creating policies that allow non-evaluation professionals to opt out of data collection or analysis so they won't have to participate in critiquing products if they are concerned it will impact their professional relationships.

When collecting data from fellow professionals, think about:

- Making their data confidential so that other project stakeholders cannot use the responses to make judgments about how to distribute materials or make decisions that fall outside of the evaluation scope.
- Creating protocols that include telling subjects how their data will be used and protected so that they will feel they can provide honest feedback without repercussions.

If you need to approach the same institutions or individuals multiple times over the course of a project, try:

- Establishing data collection and management plans with the endorsement of project leadership that
 - Limit the number of times you collect data by gathering information for multiple stakeholders at once,
 - Create ways to track data collection and participants so that you can avoid asking the same people or sites for data all the time, and
 - Use existing data to answer evaluation questions instead of collecting additional data.



AN IN-DEPTH LOOK:

Working to Ensure Non-Evaluators Had Human Subjects Training

When we decided to add Team-Based Inquiry (TBI) as a method for collecting data, the Evaluation workgroup was faced with a dilemma: how can we make sure that these new, nonprofessional evaluators have proper human subjects training so that our Institutional Review Board (IRB) feels comfortable allowing them to collect and handle human subjects data while also ensuring that we are not asking too much of our new data collectors?

We tried a number of different training protocols before settling on a process that seemed to work for everyone.

The first method that we tried was to only require workgroup leaders to take the full human subjects course and make any other team members who would collect or handle data read and sign shorter guidelines that were specific to NISE Network. This was seen as acceptable by our IRB because the data being collected were formative in nature, and the people who were leading the new TBI work had full human subjects training and could keep an eye on and train the additional data collectors as needed. Additionally, Network Leadership felt this was a good option because we were not

asking everyone to use their NISE Network time and budget to complete a two-hour human subjects course.

However, after a year of using this training method, our IRB decided that this level of training was not adequate because even though data were formative, practitioners were sometimes collecting data from protected populations such as fellow professionals and children. Therefore, we were now required to ensure that all people who were going to collect or handle human subjects data had to complete the full human subjects course. To reduce the burden on our Network partners, we worked with the IRB and the Network Leadership to arrive at two compromises.

First, we would carefully think about who actually needed human subjects training and only require it of those partners who were actually interacting with subjects to collect data or who were actively a part of the data analysis process. Second, these individuals would not need to complete their training immediately but, rather, would have many months to fit the training into their schedules.

In the end, requiring all people involved in the TBI process to complete human subjects training was easier than anticipated. We discovered that many of our partners had already taken the human subjects course, so they did not have to take it again for this project. For those who

did need to take the training, they saw it as an opportunity to increase their capacities as informal science educators. If we had the chance to write different IRB protocols from the beginning, we would advocate for setting the expectation that everyone who collects or handles human subjects data needs to undergo full human subjects training whether they are an evaluator who interacts with human subjects on a regular basis or a non-evaluator who is trying to incorporate a bit of formative evaluation into their regular work.

In the end, requiring all people involved in the TBI process to complete human subjects training was easier than anticipated. For those who did need to take the training, they saw it as an opportunity to increase their capacities as informal science educators.

CHAPTER 5:

Communicating Findings and Methods to Network Stakeholders and Beyond

Marta Beyer

The Challenge: How can we share the team's findings and strategies for conducting network evaluations with our various audiences and stakeholders?

Chapter Summary: Over the 10 years of the Network, the NISE Network Evaluation workgroup has sought to share not only our evaluation findings but also strategies and methods that helped us conduct evaluation across multiple sites, educational products, and professional experiences. When disseminating information about this work, we pushed ourselves to think broadly about different stakeholder groups and potential methods. The sheer size of the Network meant that, especially for internal communication, we had to adapt techniques to work for this context. This chapter depicts:

- How we communicated with internal stakeholders;
- What types of external audiences and additional dissemination options we considered; and
- Why we made the decisions we did when it came to sharing our work.

What we did

As evaluators, we know the importance of sharing findings and knowledge gained from working on a project. So our team spent considerable time grappling with how to share information that would both inform the Network and be of use to outside stakeholders. During this process, we:

- consulted many relevant resources,¹⁸
- reviewed exemplar reports from the field, and
- discussed potential dissemination ideas with our COV, external researchers working in the informal science education field, and the team as a whole.

Several factors, including the large scope of the evaluation and sizeable budget of the NISE Network, allowed our team to think much more broadly about dissemination options than is typical of evaluation projects. The sections below describe how we tried to think outside the box when approaching both internal and external dissemination possibilities.

Communicating with the Network Leadership and workgroups

As described throughout this document, the Network was deeply committed to continuous learning and improvement through the use of data and feedback. Because Network Leadership and those actively involved with product development were intent on using evaluation findings when making decisions about the direction of their work, our team was pushed to be especially responsive when communicating with internal audiences.



In order to provide internal stakeholders with the timely feedback they wanted, we were forced to consider:

- How can we share the latest evaluation findings in a responsive manner?
- How can we communicate this information to a large, dispersed group of Network decision-makers?

As mentioned in Chapter 1, to help us be responsive, we came up with the strategy of having evaluators sit on each of the workgroups. While resource and time intensive, this system allowed us to provide findings and evaluation insight on a frequent basis to those who needed it because we were present during all work group conference calls and in-person meetings. Moreover, joining the teams in this way permitted us to support these groups when they wanted to incorporate evaluation findings into their own work or dissemination efforts, which was especially the case for the Network Leadership.

One instance of when we worked closely with Network Leadership to integrate evaluation findings into their materials can be seen in the 2014 Report to Partners.¹⁹ This strategy of being deeply embedded on workgroups was a key means for ensuring that our work could inform those running the NISE Network.

Another method that allowed us to facilitate communication between internal network stakeholders and our team was to have members of Network Leadership be part of our various evaluation subgroups. As described in Chapter 1, we started this effort in Year 7. Including Network Leadership on our teams allowed them to become more familiar with our studies and be able to spread evaluation-related information to partners within the Network as well as to outside stakeholders such as National Science Foundation (NSF) program officers. By involving Network stakeholders in this way, we developed a unique strategy for sharing our work.

Beyond structuring teams in this manner, we also used a range of communication techniques to ensure that Network Leadership and workgroups had the data they needed to make decisions or to prepare for yearly reports and events such as NSF reverse site visits. In order to share the latest evaluation work, we often created memos and other short documents referring to initial findings from the ongoing studies, supplied additional information about areas of interest or potential future work, and presented updates either in person or over the phone.

Overall, it was thanks to methods such as these and to our frequent communication with workgroups and Network Leadership that our team was able to be responsive to the needs of communicating with a network of internal stakeholders.

Sharing information with network partners

Because our team also viewed NISE Network partners as an important internal audience, we worked hard to figure out ways to share evaluation-related information with them as well. Yet when it came to communicating findings with the Network's hundreds of partners distributed across many sites, the structure and size of the Network challenged our team to think beyond the common practice of giving one-off presentations to program participants.

Over the years, we ended up using a variety of techniques, including some of the existing Network dissemination channels, to provide timely, useful, and frequent information to the wider Network audience. Dissemination through these Network channels included:

- Presentations and/or active representation at over 20 of the NISE Network's large in-person annual meetings with subawardees and/or partners;
- Updates about evaluation efforts and findings in the monthly partner NanoBite e-newsletter and blog posts; and
- Links to over 45 front-end, formative, and summative reports posted online and connected with relevant educational products on the NISE Network website.

Through these efforts, our team was able to disseminate findings to NISE Network partners in several ways and have a significant presence within the Network. We were also able to thank those who had taken part in our studies, and better help everyone connected with the Network to understand why, in some cases, the work of the Network was shifting based on evaluation findings. By sharing evaluation work in such a transparent manner with partners, we

were able to highlight just how much the Network cared about evaluation and using data to help make decisions.

Considering additional external audiences

Besides considering key internal groups when making dissemination decisions, we also brainstormed about the various external stakeholder groups that might be interested in learning about our findings and methods for evaluating a national network. These conversations occurred throughout the project, but especially in preparation for the final few years of the grant when many of the dissemination efforts would culminate.

During this process, we pushed ourselves to think beyond the usual stakeholder group of outside funders and consider who else should learn about the findings and knowledge gained from this project.



In particular, we asked questions about who we typically reach and who we tend to miss as a part of our dissemination. Some of the main questions we wrestled with included:

- What information, methods, and findings do we have to share?
- Which stakeholder audiences do we feel are important to reach?
- Which stakeholder groups will be most likely to use different kinds of information?
- Who have we been reaching in our dissemination up until this point?
- Who else should we be reaching in the final years of the project?

When thinking broadly about these areas, we made the following list of potential external audiences:

- Professionals in related fields including the larger evaluation field, informal science education, and museums;
- Other large-scale networks;
- CEOs or grant writers from various nonprofits;
- State organizations, other funders, policymakers; and
- The general public.

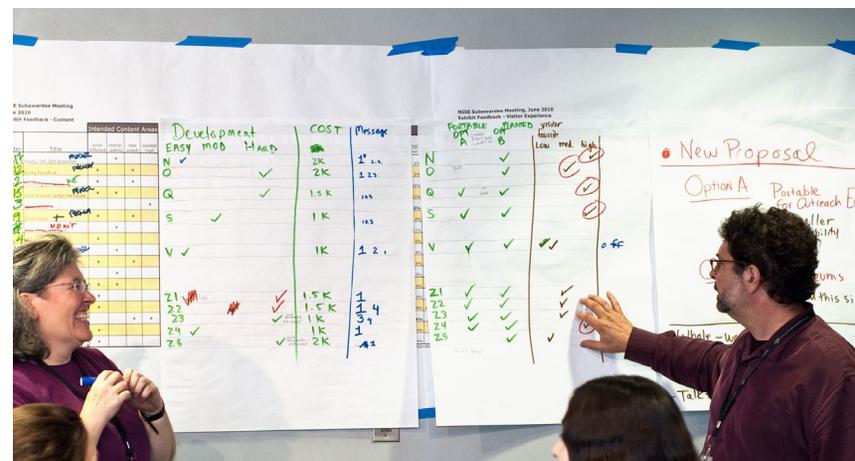
We felt that this list was different from those we had encountered before because it contained several stakeholders who are not typically considered when disseminating evaluation findings for ISE projects. Stakeholders such as nonprofit CEOs, other types of organizations, and the general public were all included because they might be interested in learning about our findings and methods for evaluating a national network. For each of these groups, our team reflected on how our work

might be relevant to them and the kinds of information that we might share. We also thought about the dissemination methods that would be most appropriate and useful. We considered the trade-offs of how much time and effort would be required to create something for each of these audiences with how crucial we felt it was to communicate with them. To learn more about our process for trying to decide if it would make sense to include the general public as a target audience, see this chapter's In-depth Look section.

While it was a valuable exercise to consider other audiences who might benefit from our work, ultimately we did not focus our efforts on all of the groups listed above. Instead, we ended up choosing the audiences we felt would be most likely to make use of our findings. Thus, besides our funders at NSF, the external audiences we targeted included professionals working in the larger evaluation field, informal science education, and museums. We decided not to focus on the other audiences because we were less familiar with their specific needs and not as convinced that they would utilize our work. Beyond these factors, the capacities and interests of our staff also figured into what groups we would disseminate to.

Prioritizing dissemination methods

In addition to thinking seriously about potential audiences beyond our internal stakeholders, we also spent time considering how best to communicate with them. Like others in the evaluation field, our team believed in matching dissemination methods to the needs of each audience.¹⁸ We felt that in order for our work to be useful, the various dissemination formats had to align with how each audience preferred to receive information. We also aimed for dissemination strategies that could convey the various types of content we wanted to share. As noted in earlier chapters, the Network context



challenged us to come up with new processes or adapt methods for evaluating the Network. And so, unlike some projects, our disseminated content included not only our study findings but also information about our COSTS and methods.

When considering what might be the most appropriate methods for disseminating our work to external audiences, we specifically asked ourselves about the techniques we had used in the past and about other areas we should explore. In repeated conversations, we kept coming back to the following questions:

- What dissemination formats have we used to date?
- What other types of dissemination techniques should we employ?
- What is the purpose of each method and which target audience would it serve?
- What strategies make sense for communicating the type of information we want to share (findings, process, etc.)?
- For each option, what types of resources are needed?

While brainstorming as a team, we considered the following dissemination methods for communicating with external audiences:

- Traditional dissemination techniques such as reports, conference presentations, and journal publications;
- Interactive options such as blogs, social media, interactive websites, webinars, and wikis;
- Reference documents including white papers, articles in popular magazines for the ISE field, or a book of reflections;
- Practitioner tools including PowerPoint slides and how-to guides; and
- Marketing approaches like press releases or newspaper write-ups.



Throughout the process of determining appropriate dissemination strategies, we remained committed to using a variety of methods and matching them with our audiences' preferences for receiving information. To do this, we considered the needs of our audiences in several ways and the types of information we were sharing. For example, at one point, to better understand where professionals in the ISE field go to learn new information for engaging the public in science, our team asked NISE Network partners on our Annual Partner Survey what journals or popular press publications they referenced as a part of their education efforts. We learned from this survey that these professionals predominantly look to *ASTC Dimensions* and *Science Magazine* rather than other publications.²⁰ This finding pushed us to continue to pursue presenting our work in both academic and popular publications within the field.

Besides thinking about our audiences' preferences, we also took into account several factors related to our team's capacities and interests when deciding which dissemination options to consider. We realized that some of the more nontraditional methods would require additional effort and experimentation beyond where we wanted to focus our efforts. Moreover, we agreed that it was unrealistic to focus on methods that required lots of upkeep, such as an interactive website, because we knew that our group would disband after NISE Network funding ended. We also recognized that even with our large budget and scope, it would be impractical to devote significant resources to a single dissemination strategy because not only did we want to utilize a wide range of methods, but we were still in the midst of ongoing evaluation work and aware that future work could arise.

The process of reflecting on our needs as well as those of our audiences helped us prioritize dissemination strategies. And

although we did not undertake every method we considered, our efforts included a mix of strategies. Broadly, our dissemination involved:

- Traditional communication techniques such as various types of evaluation reports, approachable short documents summarizing our main findings, and conference presentations over multiple years;
- Interactive options like blogs and webinars covering different aspects of our work; and
- This reference document, which gave us the opportunity to reflect on our experience, record some of our key decisions, and share lessons learned about evaluating a large-scale network with those who may face similar challenges.

To learn more about our dissemination work and see examples of our reports, visit our website: http://www.nisenet.org/About_Evaluation_Research.

What this means for other projects:

When communicating information with internal stakeholders, remember that:

- Employing a range of methods and strategies will help you be responsive to project decision-makers' needs.
- Using established project dissemination channels can facilitate the sharing of evaluation updates with project participants.

When making decisions about which external audiences to share evaluation-related information with, think about:

- Reaching out to organizations or professionals who may not be familiar with your work but could benefit from this information.
- Focusing your efforts on audiences that your team is interested in serving and who you think will actually utilize your work.

When deciding on different dissemination strategies, consider:

- Conveying information beyond findings, such as methods or aspects of your process, because your team probably has other interesting areas of work to share as well.
- Using techniques that are appropriate for the audiences identified so that your work resonates with them.
- Recognizing how your capacity and interests affect your ability to disseminate your work.



AN IN-DEPTH LOOK:

Considering the General Public for Dissemination

When thinking about who to reach with our dissemination efforts, the Evaluation workgroup considered the general public as a potential audience.

This audience came to mind for several reasons. Not only is the NISE Network project funded through a federal National Science Foundation grant, but the Network actively works to enhance the public's awareness and knowledge of nano through the development of public educational products. Thus, we thought our findings might be of interest to the public because they speak to how the general public has been impacted in these areas by the NISE Network. Furthermore, although the evaluation field agrees that it is important to share findings with public stakeholders, most evaluation work is disseminated in reports and official memos seen only by staff involved in projects or funding

sources. These factors made us realize that we should try to broaden our efforts to include disseminating information to the general public.

As we started to contemplate reaching out to the general public with our results, several questions arose: we wondered what would be the best way to communicate with this audience, and what would they be interested in learning? To try to answer these questions, our team performed a Team-Based Inquiry study, looking for examples of dissemination methods in relevant literature and asking colleagues in the field for instances of when they tried to communicate findings to nonprofessional audiences and members of the general

public. The literature provided few concrete examples of how to carry this out, and only a small number of colleagues mentioned instances of how their institutions had tried to relay evaluation information to the general public.

Even though our search did not turn up much information about how to reach this audience, we considered creating an infographic with summative evaluation findings from NanoDays, one of the Network's largest public initiatives, and adding this resource to the physical kit of NanoDays materials (for more information about NanoDays, see Chapter 3). Although not a new technique for reaching a public audience, we thought an infographic might

work well for our purposes because it could be added to the suite of posters already included in the NanoDays kit. We also knew the kit was one of the most direct ways that the NISE Network provides information to partners. By sharing findings about how others have been impacted by NanoDays in this manner, we felt current visitors and even volunteers could be more reflective about what they gain from participating in the event. We also thought an infographic could be a way to thank those who had participated in earlier data collection. In addition, if we had the opportunity to ask visitors about the infographic and hear what was most interesting and relevant to them, we could learn how to improve future dissemination work.

Yet, as we started developing an infographic for the NanoDays kits, this work raised questions for the Network staff who implement NanoDays and organize the kit creation process. Like us, they had questions about how the general public would engage with the infographic and what they would get out of it. Additionally, they had concerns about how the messages in our infographic would relate to the other NISE Network educational kit products, how the information would be

presented, and if the material would have any relevancy for the museum visitors. Finally, they had concerns that we did not have time to gather feedback from the public about the infographic and make changes before it needed to go into the kit.

After conversations with Network staff, we realized that moving forward with this dissemination strategy did not make sense at the current time because there were still questions about the overall purpose, content, and format of the infographic. Instead of disseminating findings directly to the general public, we added some relevant evaluation findings to a PowerPoint meant to help train NanoDays volunteers. We all agreed that this audience might benefit more than the general public from hearing about what we learned.

While we did not implement any dissemination with the general public, this example shows how we did share findings with NanoDays volunteers, who themselves are a distinct audience because their role falls between that of museum professional

and general visitor. What is more, we feel our story provides insight into the type of questions that may arise from both an evaluation and project management perspective when trying to disseminate findings to the public. For instance, if you

As we started to contemplate reaching out to the general public with our results, several questions arose: we wondered what would be the best way to communicate with this audience, and what would they be interested in learning?

are attempting to share information with this audience, remember to consider what methods are most appropriate for your situation and what the public will gain from this information. Also take into account how you can work with project leaders to ensure that the evaluation findings are accurately portrayed while at the same time complementing any project messages that are being communicated.

Conclusion

Marta Beyer

Throughout this document, we've tried to describe several of the issues our team faced when managing the evaluation of the largest informal education network ever formed, the NISE Network. As you've seen, we encountered a whole range of challenges including the sheer size, multi-year time frame, content, and open-source philosophy of the NISE Network. These challenges meant we had to be flexible and thoughtful in figuring out how to establish an evaluation that could react to the Network's needs including what the team setup should look like, how to define and measure impacts, and how to collect and measure data on a national scale. Besides these considerations, we also needed to ensure that data were gathered in an ethical manner, and we had to think about how to share our findings and process with those in the Network and beyond.

We hope that from reading sections, or even the whole document, you've gained a sense of how we approached these issues and the various factors we took into account when making decisions. The Introduction pointed out that this document provides examples of what worked and what didn't for our team. We know that every project is going to have different needs and that our choices won't necessarily be appropriate for you, but hopefully learning about our evaluation process provided insights that can inform your efforts.

As always, members of the NISE Net Evaluation workgroup are happy to talk further about our work and how you might take what we've learned and apply it. You can find many of our reports and guides on the NISE Net website (http://www.nisenet.org/About_Evaluation_Research). You can also reach us directly through email:

- Elizabeth Kunz Kollmann (ekollmann@mos.org)
- Marjorie Bequette (mbequette@smm.org)
- Marcie Benne (mbenne@omsi.edu)

We asked a few people from the field to consider how this document might be applicable to them. It is our hope that sharing their reactions and thoughts will give you a better sense of how this document might be useful in your own work.

“I think the document will be relevant to the Denver Evaluation Network. Now that we are sustaining on our own, how do we consider continuous evaluation of the network through a team model such as NISE Net? [One] of the biggest take aways for me in this report [was hearing about] when the excitement/honeymoon phase is over and the reality sets in—that things needed to be in place that never were... It’s true of so many networks; managing it with so many partners like you did—impressive! I really like the charts of what you did, what worked, and what you wouldn’t do again—sage advice for the rest of us clearly laid out! Addressing that different dissemination methods work better for different audiences is always an important message to hear as well.”

– Marley Steele-Inama

Director of Audience Research and Evaluation, Denver Zoo

“Activating evaluation across any network, and certainly one as large as NISE Network, is undoubtedly challenging and, as with any evaluation, there are limitations. However, this reflection document shows that integrating evaluation across a network is possible and fruitful when lead evaluators are persistently reflective and adaptive. This document is a wonderful reflection on the evaluation process in general. Doing evaluation isn’t easy, and in discussing the many challenges this team of evaluators faced, this document shows the thoughtfulness required when approaching any evaluation... I love that this is a reflection document. Evaluators are so used to helping others reflect on their work that it is nice to see a formal reflection by evaluators on their work. I think (hope) most evaluators are reflective on their practice in whatever means is fitting (e.g., a simple moment to think or organized conversation), but having a document that shows that evaluators practice what we preach (e.g., reflection) is important both for documentation purposes and knowledge sharing.”

– Amanda Krantz

Senior Associate, Randi Korn & Associates

“This document is an excellent overview of an expansive study. The authors have simplified a complex and iterative process into salient takeaways that they present clearly and succinctly. They focus on parts of the evaluation and research processes that are worthy of deeper discussion, such as ethical issues, pilot testing, and juggling multiple stakeholders. I appreciated that the authors framed suggestions and reflections with the next generation of users in mind—it made the document approachable and friendly. Additionally, the document is equally valuable as a whole and in its individual chapters. I look forward to using lessons from this reflection in my own work!”

– Kathayoon Khalil

Principal Evaluator, Seattle Aquarium

“The NISE Net Evaluation Reflection Document is a refreshing look into the history and the struggles of implementing evaluation for this massive project in an easy-to-follow account, complete with helpful tips and chapter summaries. My institution has just opened a new facility called CREATE at Arizona Science Center of which we are in the early stages of developing our own evaluation processes. From a personal standpoint, it has a similar feel to the beginnings of the NISE Net and it was incredibly helpful to read the accounts of the trials and tribulations of NISE Net's evaluation team and how they went about tackling the issues that came with such an expansive project. Particularly helpful was the process by which the annual partner survey was developed and how to collect data from fellow professionals to use for such surveys. It is so often the case that multiple departments within an institution have questions they need answered by the public and then the question often becomes, ‘How do we satisfy our institutional needs without inundating our visitors with surveys?’ As we begin to think about the dissemination of our CREATE content, it is also important that we start thinking about communicating the results of the content evaluation as outlined in [this document]. From small projects to large, anyone who is trying to start a new project involving evaluation will find this document insightful.”

– Rei Cameron

Senior Manager of the Artistry Hub, CREATE at Arizona Science Center

Glossary of Common Terms

The list below includes terms commonly used within the reflection document that have specific meanings within this report and for the NISE Network.

Committee of Visitors (COV) A Committee of Visitors is an external group of experts who reviews and provides recommendations about a piece of work to improve its performance. COVs are commonly used when a research or evaluation workgroup wants to ensure a constant outsider lens is reviewing a study and providing a broader perspective on a study's elements.

Evaluation phases Museum evaluation is often described as having three phases:

- **Front-end evaluation** is used during the early stages of a project to learn about visitors' familiarity with a topic, their interests and feelings, and what understandings they will bring to the experience.
- **Formative evaluation** is used during the design and development stage of a project to help achieve desired goals; formative evaluation involves iterative testing and modification of prototype exhibits, materials, and programs.
- **Summative evaluation** is used at the end of a project to assess the impact of a completed project; summative evaluation involves gathering data about deliverables such as exhibits and programs in order to understand the impact of those deliverables on the intended audience.

External evaluators External evaluators are often contractors for a project who are “not directly involved in the development or operation of the system being evaluated”²¹ and who are not employed by the institution creating and operating the project. In the case of the NISE Network,

external evaluators included Inverness Research Associates and Multimedia Research.

Internal evaluators Internal evaluators are usually employed directly by the institution developing and coordinating a project and involved in performing evaluations for that institution. The internal evaluation departments at the Exploratorium, Oregon Museum of Science and Industry, Science Museum of Minnesota, and Museum of Science, Boston all worked on the NISE Network evaluation at some point during its 10 years.

Nano As used by the NISE Network, “nano” referred to nanoscale science, engineering, and technology content.

NISE Network The NISE Network was the Nanoscale Informal Science Education Network composed of over 500 informal science education institutions and funded through two National Science Foundation grants totaling over \$40 million.

Network Leadership The Network Leadership group was composed of individuals who oversaw the leadership and day-to-day operation of the NISE Network including the project principal investigators, operational group, and the leader of the Evaluation workgroup. During the 10 years of the Network, individuals within this group came from the Exploratorium, Science Museum of Minnesota, Sciencenter, Museum of Life and Science, and Museum of Science, Boston.

Partners Partners were the institutions and individuals who were involved in the Network by developing or somehow using NISE Network products and practices with the public or participating in Network professional development activities. By 2015, the NISE Network comprised over 500 partner institutions and over 1,000 individuals.

Professional Impacts Professional Impacts was a term used by the Network to refer to the goals that the NISE Network had for the informal science education and university professionals

who participated in the Network. You can find a complete copy of the professional impact goals in Appendix B.

Public Impacts Public Impacts was a term used by the Network to refer to the goals that the NISE Network had for members of the public who took part in Network education activities. You can find a copy of the NISE Network content map, that covers learning goals in Appendix A. Additionally, the NISE Network had public goals related to the reach of public products and impacts on public interest in and understanding of the relevance of nano.

Stakeholders Stakeholders are the groups for whom a project holds some kind of interest. Primary stakeholders for the NISE Network Evaluation workgroup included Network Leadership, workgroups, Network partners, and the National Science Foundation. Secondary stakeholders included the informal science education field, museum field, and evaluators.

Workgroup (often referred to as “team”) NISE Network workgroups were cross-institutional teams that organized the project work and developed products and practices. Over the 10 years of the Network, there were workgroups and ad hoc planning groups related to: Network Administration, Annual Meeting Planning, Content Steering, Community, Exhibits, Forums, Inclusive Audiences, Nanoscale Education Outreach, NanoDays, Nano and Society, Network Media, Online Brown-Bags, Programs, Research Center - Informal Science Educator Partnerships (RISE), Visualization Lab, and Website.

Notes

- 1 Svarovsky, G., Goss, J., & Kollmann, E.K. (2015). Public reach estimations for the NISE Network. Notre Dame, IN: NISE Network.
- 2 Throughout Years 4–10 of the NISE Net, the primary institutions involved in the evaluation were the Museum of Science, Boston (MOS), Oregon Museum of Science and Industry (OMSI), and Science Museum of Minnesota (SMM). However, in Year 10, the University of Notre Dame also became involved in the evaluation when a key workgroup member from SMM moved there.
- 3 Reich, C., Ellenbogen, K., & Benne, M. (2009). *NISE Net public impacts summative evaluation progress on Year 4 activities*. Retrieved from http://www.nisenet.org/catalog/evaluation/year_4_summative_evaluation_progress_memo
- 4 OMSI was added to the Evaluation workgroup because OMSI, as a museum, was becoming more deeply involved in the Network, and they had an existing evaluation department with capacity to take on work. At the same time, the Exploratorium was becoming less involved in the NISE Network, so they decided to become less involved in the evaluation.
- 5 Pattison, S., Cohn, S., & Kollmann, L. (2014). *Team-Based Inquiry: A practical guide for using evaluation to improve informal educational experiences*. (2nd ed.). Retrieved from http://www.nisenet.org/catalog/tools_guides/team-based_inquiry_guide
- 6 Herr, K., & Anderson, G. (2005). *The action research dissertation*. Thousand Oaks, CA: Sage Publications. Loucks-Horsley, S., Stiles, K., Mundry, S., Love, N., & Hewson, P. (2010). *Designing professional development for teachers of science and mathematics* (3rd ed.). Thousand Oaks, CA: Corwin Press.
- 7 Cochran-Smith, M., & Lytle, S. (2009). *Inquiry as stance*. New York, NY: Teachers College Press.
- 8 Preskill, H., & Torres, R. (1999). *Evaluative inquiry for learning in organizations*. Thousand Oaks, CA: Sage Publications. King, J. (1998). Making sense of participatory evaluation practice. *New Directions for Evaluation*, 80, 57-67.
- 9 Cousins, B., Goh, S., Clark, S., & Lee, L. (2004). Integrating evaluative inquiry into the organizational culture: A review and synthesis of the knowledge base. *The Canadian Journal of Program Evaluation*, 19(2), 99-141. Huffman, D., Thomas, K., & Lawrenz, F. (2008). A collaborative immersion approach to evaluation capacity building. *American Journal of Evaluation*, 29(3), 358-368. King, J., & Volkov, B. (2005). A framework for building evaluation capacity based on the experiences of three organizations. *CURA Reporter*, 35(3), 10-16.
- 10 Svarovsky, G., Goss, J., Ostgaard, G., Reyes, N., Cahill, C., Auster, R., et al. (2013). *Summative study of the Nano mini-exhibition*. Retrieved from http://www.nisenet.org/catalog/evaluation/public_impacts_mini-exhibition_study_year_8_summative_evaluation
- 11 Center for Universal Design, College of Design, North Carolina State University. (2008). *About UD*. Retrieved from: http://www.ncsu.edu/ncsu/design/cud/about_ud/about_ud.htm
- 12 American Evaluation Association. (2011). *American evaluation association public statement on cultural competence in evaluation*. Retrieved from <http://www.eval.org/p/cm/ld/fid=92>
- 13 U.S. Food and Drug Administration. (2014). *Institutional Review Boards Frequently Asked Questions-Information Sheet*. Retrieved from <http://www.fda.gov/RegulatoryInformation/Guidances/ucm126420.htm>
- 14 American Evaluation Association (2004). *American evaluation association guiding principles for evaluators*. Retrieved from <http://www.eval.org/p/cm/ld/fid=51>
- 15 National Institute of Health, Office of Extramural Research. (n.d.). *Protecting human research participants*. Retrieved from <http://phrp.nihtraining.com/users/login.php?l=3>
- 16 NISE Network. (2012). *NISE Net guidelines for collecting and handling data*. Retrieved from http://www.nisenet.org/catalog/tools_guides/nise_net_guidelines_collecting_handling_data
- 17 King, J. A., Lawrenz, F., Reich, C., Bequette, M., Pattison, S., Nelson, A. G., et al. (2015). *Complex adaptive systems as a model for network evaluations (CASNET): Grounding and Tier 1 groups case study*. Saint Paul, MN: University of Minnesota.
- 18 Patton, M. Q. (2008). *Utilization-focused evaluation*. (4th ed.). Los Angeles: SAGE Publications. Torres, R.T., Preskill, H., & Piontek, M.E. (2005). *Evaluation strategies for communicating and reporting: Enhancing learning in organizations*. (2nd ed.). SAGE Publications. Retrieved from <http://www.sagepub.com/books/Book225545>. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. (2013). *Evaluation reporting: A guide to help ensure use of evaluation findings*. Retrieved from http://www.cdc.gov/dhds/p/docs/Evaluation_Reporting_Guide.pdf. Weiss, C. H. (1998). *Evaluation*. (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- 19 NISE Network. (2014) *Report to partners*. Retrieved from: <http://nisenet.org/catalog/nise-network-impacts-report-partners>
- 20 2013 Annual Partner Survey. *Context document and summary results from close-ended questions*. [Data file.] Portland, OR: NISE Network.
- 21 The Robert Gordon University, Center for the Enhancement of Learning & Teaching. (1998). *Tertiary level teaching: Evaluating the effectiveness of the teaching/learning process—external and internal evaluation*. Retrieved from: <http://www2.rgu.ac.uk/celt/pgcertlt/evaluating/eval4.htm>

Appendix A: Content Map

<http://www.nisenet.org/catalog/engaging-public-nano-key-concepts>

To begin to understand nano, we can explore four main concepts.

1.

Nano is small and different

Nanoscale things are very small, and often behave differently than larger things do.

2.

Nano is studying and making tiny things

Scientists and engineers have formed the interdisciplinary field of nanotechnology by investigating properties and manipulating matter at the nanoscale.

3.

Nano is new technologies

Nanoscale science, engineering, and technology lead to new knowledge and innovations that weren't possible before.

4.

Nano is part of our society and our future

Nanotechnologies have costs, risks, and benefits that affect our lives in ways we cannot always predict.

NISE Network Goals for Professional Development

The NISE Network is a community that aims to increase the capacity of the informal science education field to provide nanoscale science, engineering, and technology educational experiences to diverse public audiences. The fundamental purpose of the Network is to raise the level of public awareness and understanding of this emerging field of research. The Network provides different ways for partners to participate, appropriate to each organization's mission, capacity, and audience.

Overarching goal: Increase the **readiness of individual practitioners and the capacity of the field** of informal science education (ISE) to foster public awareness, understanding, and engagement with nanoscale science, engineering, and technology and its relationship with our lives, society, and environment (“nano”).

As a result of participating in NISE Network professional development activities, professionals will:

1. Identify with a **broader community** that includes scientists and museums
 - Short-term: Professionals value networking opportunities offered by NISE Network.
 - Short- to medium-term: Professionals value participation in the Network and the opportunities for collaboration the NISE Network offers.
 - Long-term: It is a norm in the ISE field to collaborate with other organizations.
2. Value local **research-ISE collaborations**
 - Short- to medium-term: Researchers and ISE professionals begin to collaborate on discrete nano-related projects.
 - Long-term: Research and ISE organizations create strong and lasting partnerships.
3. Understand and appreciate **key concepts** in nanoscale science, engineering, and technology and its relationship with our lives, society, and environment
 - Short-term: Professionals are aware of nano concepts.
 - Short- to medium-term: Professionals understand nano concepts.
 - Short- to medium-term: Professionals are enthusiastic about engaging their public audiences in nano.
 - Long-term: It is the norm in the ISE field to engage diverse public audiences in nano.
4. Understand **theories, methods, and practices** for effectively engaging diverse public audiences in nano
 - Short-term: Professionals are aware of theories of learning, educational methods, and effective practices for engaging the public in nano.
 - Medium- to long-term: Professionals apply theories of learning, educational methods, and effective practices when engaging the public in nano.
5. Utilize **professional resources and educational products** for engaging diverse public audiences in nano
 - Short-term: Professionals are aware of professional resources and public educational products for engaging the public in nano.
 - Short- to medium-term: Professionals have the tools, skills, and confidence to use, adapt, and create educational products for engaging the public in nano.
 - Long-term: Organizations integrate nano into ongoing ISE efforts.

NISE Network Years 6–10 Logic Model

OVERARCHING NETWORK GOALS

1. In partnership with the research community, develop the necessary capacities and resources to achieve a widespread, sustainable impact on the ISE field.

2. Engage the development and delivery power of the network community to raise the level of public awareness, engagement, and understanding of nanoscale science, engineering, and technology.

