



Summative Evaluation of the Northeast Big Data Innovation Hub

Building a Community- Responsive Network

February 12, 2021

Rebecca Joy Norlander, Uduak Grace Thomas, Nezam Ardalan,
Kathryn Nock, & Nicole LaMarca



© 2021 by Knology Ltd., under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND).

Recommended Citation

Norlander, R. J., Thomas, U. G., Ardalán, N., Nock, K., & LaMarca, N. (2021). Summative Evaluation of the Northeast Big Data Innovation Hub: Building a Community-Responsive Network. Knology Publication #NSF.159.560.02. Knology.

Date of Publication

February 15, 2021

Prepared for

Florence Hudson

Executive Director

Northeast Big Data Innovation Hub,

Columbia University

475 Riverside Dr., Interchurch Center – Room 320-H

New York, NY 10115



This material is based upon work supported by the National Science Foundation under Grant No. (NSF #1946615, #1946558, and #1947000). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.



Knology produces practical social science for a better world.

tel: (347) 766-3399
40 Exchange Pl. Suite 1403
New York, NY 10005

tel: (442) 222-8814
3630 Ocean Ranch Blvd.
Oceanside, CA 92056

Knology Publication #

NSF.159.560.02



Executive Summary

Knology is leading independent evaluation of various efforts led by the Northeast Big Data Innovation Hub, referred to as the “Northeast Hub” or the “Hub” in this report. The evaluation explored the impacts of programmatic activities that are intended to increase stakeholder engagement in big data science, develop shared resources and services, and support data science education and workforce development.

To assess these impacts, the evaluation featured two rounds of collaborator interviews conducted approximately seven months apart. For the first round of interviews, Knology interviewed two distinct groups of stakeholders: **Project participants** are data science researchers and practitioners whose role was explicitly described in the grant proposal to the National Science Foundation (NSF), the funding mechanism supporting the Hub; and **Community participants**, who were not listed in the NSF grant proposal, but have engaged with the Hub in a variety of ways and to varying degrees. To better understand more recent collaborations such as those focused on the COVID-19 pandemic, Knology conducted a second round of interviews with Hub participants. For this round of interviews, the Northeast Hub planning team selected data science researchers and practitioners who are actively involved in the Northeast Hub’s current projects.

Across both interview rounds, Knology conducted interviews and analyzed data through the lens of five specific measures that were designed to capture the Hub’s progress towards its strategic goals. Generally speaking, these measures assess the Hub’s ability to reach various audiences, provide a forum for data science-focused partnerships and collaborations, the benefits of being affiliated with the Hub, the degree to which its activities support the field’s projects and goals, and how its work is shaping the field.

In terms of the first of these measures, **Reach / Engagement**, the first round of interviews indicated that participants have varying levels of connection and participation with the Hub. We found evidence that the Hub has convened stakeholders across multiple sectors including academia, industry, nonprofit, and government. It has also provided a forum for data science researchers and practitioners to discuss topics of interest and forge collaborations, some of which have led to new funding. This ties into the Hub’s efforts to support the **Integration / Interconnectedness** of the community, which is the second evaluation measure. Evidence from the first round of interviews showed that the Hub effectively facilitates cross-sector networking and collaboration, with potential applications to societal challenges.

Results from the second round of interviews, which were done in December 2020 and January 2021, supported and expanded on the findings from the first round on both measures. Respondents thought that communications better showcase the Hub’s work and the value it provides, particularly through a recent website redesign and revised strategic plan. There’s also evidence that the Hub is making a concerted effort to build a more diverse pool of stakeholders, including an increased focus on student participation. Fostering and supporting community building is undeniably an area of strength for the Northeast Hub.

While affiliation with the Hub helped to open doors for some participants, the first round of interviews indicated that some members of the Hub felt less connected to its efforts than they

would like, in part due to a lack of awareness about the Hub's active projects. This was highlighted not as criticism, but as a call for the Hub to streamline its efforts in an attempt to achieve both **Impact** and **Scale**.

Respondents in the second round of interviews were more satisfied with their relationship to the Hub. It was clear that the Hub brings unique value to the data science community that is irreplaceable, and that its loss would be detrimental to the field. Conversation around scale during the second round of interviews was contemplative, acknowledging the potential for growth but with considerations about how best to support larger and bigger projects given the Hub's current capacity. As evaluators, we agree that the Hub's recent focus of amplifying the work of stakeholders is an effective approach to demonstrating both impact and scale.

On the topic of **Representativeness**, the first round of interviews suggested that there is an alignment between the Hub's activities and interests and those of its stakeholders. Harkening back to evidence of the Hub's unique position in the data science community, participants observed that projects aligned with the Hub have greater visibility, and a sense of greater credibility particularly when approaching potential partners in other sectors. Second round interviewees echoed many of these same sentiments, in addition to feeling overwhelmingly supported by the current Hub staff. Critical to long-term engagement and participation is a sense of alignment with the Hub's mission. That, coupled with clear investment in the community in terms of resources and support, will go a long way towards ensuring its sustainability.

Overall, it is clear that Northeast Hub leadership is committed to building and supporting a thriving data science community. Based on the feedback from participants that Knology received throughout the evaluation, that dedication is undeniably yielding fruit. Recognizing that people in the field can provide the best insights into what they need, Hub staff have worked to center the community's ideas and promote community-led initiatives that draw broad interest and participation. Given the number of perspectives and potential projects, it is important to strike balance between being inclusive and maintaining a manageable portfolio. Feedback from the interviews suggests that the Northeast Hub has been able to successfully navigate that balance.



Table of Contents

Executive Summary	i
Introduction	1
Background	1
Evaluation	3
Evaluation: Round One	5
Methods	5
Participants	5
Instrument	6
Analysis	6
Results	6
The Hub Demonstrates Reach and Engagement	6
The Hub as a Unique Force for Integration & Interconnection	8
The Impact of the Hub on Stakeholders	10
The Hub is Aligned with the Data Science Community	11
The Hub is Positioned for Long-Term Scale	12
Discussion & Recommendations	13
Recommendations	15
Response to Evaluation: Round One	17
Evaluation: Round Two	23
Methods	23
Participants	23
Instrument	23
Analysis	23
Results	23
Recent Progress: Reach & Engagement	23
Recent Progress: Integration & Interconnection	25
Recent Progress: Impact	27
Recent Progress: Representativeness	29
Recent Progress: Scale	30
Discussion & Recommendations	32
Recommendations	33

References	34
Appendices	35
Appendix A: Round One Interview Protocols	35
Round 1 Project Team Interview Instrument	35
Round 1 Community Partner Interview Instrument	37
Appendix B: Round Two Interview Protocol	39
Round 2 Interview Instrument	39



Introduction

The mission of the Northeast Big Data Innovation Hub (“Northeast Hub” or “Hub”) is to build and strengthen partnerships across industry, academia, nonprofits, and government to address societal and scientific challenges, spur economic development, and accelerate innovation in the national big data ecosystem. Within that context, the Hub was designed to be a community convener, collaboration hub, and catalyst for data science innovation in the Northeast region.

Background

The Northeast Hub was initially conceived to help support data science knowledge and application, serving as a way to coalesce the efforts of researchers and practitioners working with large datasets. In June 2019, following its initial founding as part of the National Science Foundation (NSF) Big Data Spokes program, the Northeast Hub received funding (NSF award #1916585) along with Hubs in the South, Midwest, and West regions. In addition to supporting core activities of the Northeast, South, Midwest, and West Hubs, NSF also required the Hubs used their funding to build a strategic plan¹, create seed funding opportunities, and design a National Coordination Committee across the four Hubs.

The Northeast region includes project partners working primarily in Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont. Housed at Columbia University, the Northeast Hub is a collaborative entity guided by many voices in order to sustain impact across diverse societal issues. The project team (PI, Co-PIs, Executive Director, Project Manager) has been supported by a Steering Committee, Advisory Board, Seed Fund Steering Committee, Project Activity Leaders and the National Coordination Committee. Since its launch, the Hub has engaged in many activities aimed at growing and supporting the data science community.

To address local issues in collaboration with cross-sector communities across the Northeast, the Hub initially structured work within four application areas and four cross-cutting themes:

Application Areas

- Education
- Health
- Rural / Urban Spectrum
- Science

¹ The strategic plan for the Northeast Hub is available at <http://nebigdatahub.org/wp-content/uploads/2020/06/NEBDHub-Strategic-Plan-6.1.2020.pdf>

Cross-Cutting Themes

- Data Literacy
- Data Sharing
- Responsible Data
- Privacy & Security

Each application area, as well as each cross-cutting theme, was designed to have a Volunteer lead from the community. Additionally, twelve projects that focused on these specific application areas and themes were highlighted as part of the proposal. Each project had a specific team named in the proposal.

Coinciding with the onboarding of Florence Hudson as the new Executive Director in May 2020, the Hub reassessed these priority areas, opting to focus on four core focus areas to guide the Hub's use of its resources and maximize its impact on the data science community. These redefined core areas are of **Education + Data Literacy, Health, Urban to Rural Communities,** and **Responsible Data Science: Security + Privacy + Ethics.** Details on the objectives for each focus are provided in the Hub's strategic plan.

Conversations with the Hub staff indicate that building an inclusive data science community is a priority for Florence Hudson and her team. Students in particular have been a key demographic that the Hub has attempted to draw in through providing research experiences for undergraduates, encouraging them to contribute, and implementing their suggestions. A recent example illustrating this commitment is the launch of a student paper challenge to encourage students to share their work with the community. Hub leadership have also intentionally cast a wider net with their outreach efforts, particularly towards traditionally marginalized groups and smaller schools that lack data science programs. They have also invested in building networks that connect individuals and institutions with shared interests to explore opportunities for collaboration. Moving forward, the Hub intends to continue supporting efforts under its core areas through various seed projects.

The Hub also recently updated its website (<https://nebigdatahub.org/>) to make it more inviting and user-friendly, better support the data science community, and provide clear opportunities for involvement. Updates include a simplified portrayal of the Hub's current work, and links to detailed descriptions of and content from the different projects under its core areas. One of these is the COVID Information Commons, which aims to facilitate knowledge sharing and support collaboration across various COVID research efforts. Another example is the Northeast Student Data Corps, which is a community-developed initiative to improve data literacy by teaching data science fundamentals to students in the northeastern United States, particularly those in under-resourced institutions. In addition, the Hub is communicating with the data science community through channels such as Twitter, LinkedIn, YouTube, and email newsletters.

The Northeast Hub's website also offers an option for members of the data science community to submit new projects under each of the four focus areas, which may be added based on available Hub and community resources. Future plans for the website include combining resources from each project to target specific audiences. This could include a list of resources for students interested in learning about data science, and for educators who are interested in establishing data science courses or programs at their institutions.

Evaluation

Knology is leading independent evaluation of the Northeast Hub's activities. The evaluation is designed to assess the progress the Hub has made in relation to its four strategic goals:

1. Build collaborations to address real-world challenges through translational data science approaches;
2. Foster innovation and scale endeavors that reflect regional interests and align with national priorities related to data science;
3. Support and promote representative community engagement / impact across all Hub activities; and
4. Increase data science capacity and talent, emphasizing underserved communities.

The evaluation has progressed along five measures:

- **Reach / Engagement** measures the number and variety of communities affected by Hub activities and the number and variety of partnerships that the Hub works with;
- **Integration or Interconnectedness** measures the extent to which Hub partnerships amplify and cross-pollinate Hub-related activities to create economies of scale;
- **Impact** measures the perceived added benefit for stakeholders related to collaboration or integration with the Hubs, and the degree to which those who volunteer their time on Hub activities are satisfied with those efforts;
- **Representativeness** is the extent to which Hub activities align with and support the goals of partner and community projects; and
- **Scale** is the extent and depth of change attributable to the activities, projects, and initiatives.

In designing the evaluation plan, Knology sought to capture the breadth and depth of the Hub's current activities, measure progress against goals to date, and identify how the Hub could expand its impact and achieve long-term sustainability. In June 2020, Knology conducted a first round of interviews with 22 data science researchers and practitioners across a range of institutions and sectors to understand the scope of the Hub's efforts and how those efforts have produced meaningful results. These interviews shed light on how the Hub works with the community to leverage data science to catalyze action from knowledge and insight and enable the development of new solutions to societal problems.

Following Knology's first round of evaluation, the Hub committed to further developing community programs and events that bridge academia and other sectors. The Hub also devised plans to reach broader and non-academic audiences – for example, by co-sponsoring events with partners and collaborators. Additionally, the Hub is building out its online presence through various avenues. As noted above, this includes highlighting data science projects, outcomes, and opportunities on the website.

To capture further progress made along the five measurement areas, Knology held a second round of interviews with 12 participants in December 2020 / January 2021. This time, Knology interviewed data science researchers and practitioners who are more actively involved with the Northeast Hub's current projects. These individuals were selected by the Northeast Hub planning team. Interviewing these stakeholders stemmed from a desire to better understand more recent collaborations – such as those around the COVID-19 pandemic – rather than revisit the Hub's historical accomplishments, which were well-documented in the first report.

This report summarizes activities from both interview rounds with the participants in the Northeast Hub. It includes individual chapters for each evaluation round that describe the methods used, results, and offer a set of recommendations based on the findings. We structured the results section using the five measurements mentioned above, while the discussion section is structured using the Hub's stated strategic goals. Throughout the report, we include quotes from the interviews that help to highlight the community's thinking about the Hub and its perceived value. Additional sections cover the Hub's responses to the evaluation of its activities and programs conducted by Knology, and the interview protocols used that were used during both rounds.



Evaluation: Round One

Methods

Participants

Knology worked with the Executive Team at the Hub to identify individuals who could speak to the impact that the Hub has had to date as well as opportunities for success moving forward. They created list of individuals from the people named in 2018 proposal to NSF either as project team members for the application areas, cross-cutting themes, or projects (Project Participants), as well as individuals whose engagement, while not originally in the grant, has been critical or evolved since the funding was awarded in 2019 (Community Participants).

Proposal Projects

- EmPAWRed Exposome Data Exchange Pilot: Reducing Barriers to Data Access and Interoperability: *Health, Rural/Urban, Data Sharing, Privacy & Security, & Responsible Data Science.*
- Smart Cities Data Exchange: *Health, Rural/Urban, Science, Data Sharing, Privacy & Security, & Responsible Data Science.*
- Seamless Licensing Agreements Collaboration: *Data Sharing, Privacy & Security, & Responsible Data Science.*
- Responsible Data Science in Government: *Rural/Urban, Science, Data Sharing, Privacy & Security, & Responsible Data Science.*
- Building Capacity with The Carpentries: *Data Literacy, & Responsible Data Science.*
- Data Science for All: *Data Literacy & Responsible Data Science.*
- Management Training: Sandra Will, Will Squared LLC: *Data Literacy, & Responsible Data Science.*
- Collaborative Resource and Understanding eXchange: *Data Literacy & Responsible Data Science.*
- Learning and Career Development Initiative: *Education & Responsible Data Science.*
- Outreach on Educational Data Mining: *Education, Data Literacy, & Responsible Data Science.*
- Cybersecurity Risk Initiative: *Privacy & Security, & Responsible Data Science.*
- Computational Steering: *Science & Responsible Data Sharing.*

A total of 22 individuals agreed to do interviews:

- **Project participants** ($n = 12$) were defined as individuals named in the NSF proposal as members of a team working on an application area, cross-cutting theme, or project;
- **Community participants** ($n = 10$) were defined as individuals who have had some connection with the Hub which was either undefined in the proposal or has evolved since.

Among the Project and Community Participants, some were also part of the various support committees outlined in the previous section. Of the 22 participants, 6 were either on the Hub's Steering Committee, the Advisory Board, or a co-PI. To preserve anonymity, we refer to all six of these individuals as members of the Hub "Leadership Team" when we note their responses in the analysis.

Instrument

The interview protocol was designed to assess progress on the five measures guiding the evaluation, listed in the introduction. It consisted of eight primary questions with secondary follow-up questions or probes as needed. Interviewers adapted the protocol to each interviewee based on their personal history and connection to the Hub. Interview protocols for Project versus Community participants varied only slightly in how they framed similar questions. See the Appendix to this report for both interview protocols used in Round 1.

Three Knology researchers participated in data collection, rotating through roles. All interviews took place over Zoom and were recorded with verbal consent. Each interview lasted for approximately 30 minutes.

Analysis

Analysis of the interview data was performed by the same researchers who conducted the interviews. We used a grounded theory approach to see what themes surfaced in the conversations organically. We used the five measures to interpret and provide structure to the analysis, as is evident from the headings in the Results section.

While the groupings of Project and Community participants were helpful for the interviewee selection process, data from both groups were largely analyzed together. Due to the broad criteria for inclusion in either group we did not anticipate meaningful analytic categories to result from the interviews. However, we used designations when knowing more about interviewees' roles provided additional insight into their comments.

Throughout this report we refer to individuals using the singular "they" to protect identity.

Results

The Hub Demonstrates Reach and Engagement

Reach / Engagement measures the number and variety of communities affected by Hub activities and the number and variety of partnerships the Hub works with. We understand this to mean how the Hub occupies a central role as a convener that has engaged stakeholders through events and communication efforts.

Our analysis resulted in the following key findings related to Reach / Engagement:

- The Hub acts as convener for a wide variety of stakeholders, resulting in tangible outputs;
- Members of the data science community have benefitted from past events and recognize the Hub's outreach efforts, but acknowledge room for improvement; and
- Stakeholders are eager for additional ways to engage with the Hub but not always sure how to do so.

Although both Community and Project participants had varying levels of engagement with the Hub, the majority of all interviewees described the Hub's role as a convener as key to building collaborations and fostering innovation. Evidence suggests that the Hub has effectively connected stakeholders across research, academic, nonprofit, industry, and – to a lesser extent – governmental sectors. One result of this convening function has been the development of

collaborative grant proposals, some of which are aimed at advancing the inclusivity of the data science community in the Northeast region. For example, one interviewee had recently submitted a proposal to work with teachers in Title I schools to bring data science education to students in under-resourced areas. The interviewee indicated various members of the Hub that were integral to the proposal project, saying, *"These relationships are very much alive, it's a network that's really important to the work we've been doing."* A Leadership Team member reiterated this idea, saying, *"I've benefitted from the ecosystem that has emerged around [the Hub]."*

Furthermore, interviewees described the Hub as having an *"umbrella role"* that unites data science activities across the Northeast region. When asked what might be lost if the Hub didn't exist, one participant emphasized that this convening function is vital. *"If the Hub didn't exist, or if it went away, then the [loss of the] opportunity to have a regional convener for activities around big data and the opportunity to work with that convener would be unfortunate."*

Interviewees, especially Project Participants who had been involved with the Hub for many years, spoke positively about Hub-organized events. One described a Hub-related convening as *"one of the best events I attended,"* and many others expressed a desire for increased involvement in Hub activities. Interviewees felt that regular events in different Northeast locations would help keep the Hub *"top of mind,"* and increase engagement though they acknowledged the prohibitive expense of in-person gatherings. One participant felt that having a talk series over Zoom could bolster the Hub's public presence. This participant also noted that now would be an appropriate time to begin virtual meetings or workshops given the current restrictions on gatherings due to COVID-19. Several other interviewees also noted the potential benefit of online convenings while simultaneously recognizing the general feeling of saturation with video conferencing. For future Hub-related events, one interviewee suggested that the Hub could provide clearer, actionable follow-up.

One respondent commended recent Hub efforts related to website redesign and others appreciated the Hub's community outreach efforts via its newsletters. However, the interview data suggests that the broader data science community may be unaware of all the Hub's activities. Some attributed their lack of awareness to the general chaos surrounding COVID-19. Participants mentioned feeling out-of-touch or only knowing about annual events, while others were aware of more frequent Hub-related activity.

The varying degrees of awareness may be related to their level of engagement with the Hub, or with the Executive Team directly. The level of connection also seemed to vary with geographic location. Generally speaking, interviewees that were based in New York City had more frequent engagement with the Hub and knew more about its efforts. This suggests that the Hub may have a stronger presence there than in some other Northeast areas. Some interviewees noted that a communications and outreach strategy aimed at reaching participants in other parts of the Northeast could help broaden engagement with the Hub. Along these lines, another participant suggested that the Hub could emphasize Spanish-speaking workshops or events as one way to be more inclusive with outreach efforts.

Overall, participants wanted the Northeast Hub to have more frequent activity, whether through online activities or through events and publications. One participant stated that it would be helpful for the Hub, *"To get back on everybody's radar on a more regular basis so ... that opportunities and engagement areas are more fully known."* Considering other opportunities for deeper engagement, one participant mentioned utilizing volunteer-led working groups to

help the community more fully engage with the Hub and to capitalize on areas of shared interest. One interviewee felt that the Hub could play a bigger role in supporting efforts to educate members of the public about big data or artificial intelligence. They noted that this could include support for learning happening in public libraries or community centers which are attended by non-academic audiences.

Another frequently mentioned obstacle to participating in Hub-related activities and events was time. This was true for both Project and Community participants, who said that although they saw the Hub as a valuable entity, finding time in their schedules to attend events and other activities was often a real hurdle. Yet even those who had the lowest levels of current engagement seemed to desire to re-engage, suggesting that they found value in the Hub and that time limitations were not insurmountable. However, they were occasionally unsure about how to become more involved, how proactive to be, or what stage in a research project was optimal for pursuing the Hub's support.

Interviewees had a number of suggestions to increase overall reach and engagement. One saw the need to "bring in new blood," noting that there has been limited infusion of new participants since the Hub's founding. Others had specific suggestions for getting more participants involved. This included things like sharing newsletters spotlighting opportunities across the field (such as hackathons or creating open source datasets), and using the Northeast Hub website to offer clear invitations to action or specific types of involvement. They acknowledged having to look across many sources for opportunities in data science and felt the Hub as a centralized communicator and resource would be extremely valuable for the field.

The Hub as a Unique Force for Integration & Interconnection

Integration or **Interconnectedness** measures the extent to which Hub partnerships amplify and cross-pollinate Hub related activities to create economies of scale. We understand this to mean ways that the Hub promotes cross-sector collaboration and works to reduce disciplinary silos. It also includes how the Hub connects academic research with real-world application. Interviewees, aware that much of this connecting function falls on the Hub's Executive Team, offered strategies to help amplify efforts while reducing burden.

Our analysis resulted in the following key findings related to Integration and Interconnectedness:

- Networking efforts of the Executive Team are recognized and appreciated by stakeholders;
- The Hub has successfully facilitated collaboration within academia as well as between academia and other sectors; and
- The Hub connects data science research with opportunities for meaningful application, a function that is particularly valuable for data science students.

By far the most frequently mentioned benefit of the Hub was related to networking, as a way of generating new ideas and pushing the boundaries of the data science field. Most participants said they had benefited from being in the Hub network or that they expected to benefit in the future.

"It's a place to meet people... There are times and places where I've run into people who I wouldn't otherwise run into, [and] that has led to ideas that would not otherwise have occurred. I think that is the principle outcome of participating in the Hub."

A similar comment confirmed the idea that the existence of the Hub serves to break down barriers that normally exist in the field of data science: *"There are a lot of really smart people in the Hub who I wouldn't naturally bump into without something like the Hub, so I think that's a real benefit."*

Evidence suggests that direct relationships with either current or past members of the Executive Team are largely responsible for the Hub's identity as a connecting force. Interviewees unambiguously praised both previous and newer members of the Executive Team and their ability to facilitate integration:

- *[Names] are a "superstar team!"*
- *"[Name] does a hell of a lot of work...they're awesome. If [the Hub] lost them they'd be in big trouble. They're a lynchpin of the organization."*
- *"[Name] is probably the best person to do [networking] because they are definitely a connector."*
- *"[Name] was really good at reaching out and saying 'Hey you're looking at this issue involving data and [another researcher] who is in health care is also looking at data. Maybe you two should talk?'"*
- *"Somebody like [Name]...to be such a supporter and really championing our work and connecting us with people, that's invaluable."*

Interviewees commented on the unique position of the Hub to facilitate connection between academic researchers who tend to operate in disciplinary silos. One interviewee who works extensively with research networks called this the *"cross-fertilization of research activity."* This function is particularly pertinent because collaboration across different institutions can be challenging, due to self-interest (in particular when it comes to recognition for research and procuring funding).

Those interviewed for this evaluation clearly identified the Hub as a body able to connect researchers within and across institutions of higher education. For example, a member of the Leadership team commented that, *"the Hub has the opportunity to bring together people who are working in data science in different research fields together in a way that probably wouldn't happen without something like the Hub. I think that is an irreplaceable role."* Another participant noted a similar idea, saying, *"The Hub enables, in a very non-threatening environment, academia to come together for a greater purpose than themselves."* Ultimately, this interconnectedness can lead to a greater capacity that is not possible when institutions work separately. As one participant noted, *"I like to talk about the Hub as taking on challenges that are above and beyond what any single institution can do alone."*

Additionally, several participants specifically highlighted the value of having a central entity that integrates diverse stakeholders from across different sectors in ways that can profoundly shape the field of data science. In the words of one interviewee,

"If the Hub didn't exist, how would we have those kinds of conversations? Who would bring together all the key players that are involved in decision making for data science, especially for data science education and training? Who would bring those people together if we did not have the Northeast Big Data Hub?"

Participants also felt that the Hub could do more to broker relationships between academia and other sectors like industry, government, and nonprofits to address real-life challenges facing society. One participant viewed the Hub as primarily an academic entity but saw its potential for impact as a connector with industry. Another agreed, noting that not all academics are interested in pursuing partnerships that focus heavily on real-world engagement, but those who do, tend to be *“self-sorting.”* Two mentioned the use of Research-Practice Partnerships (or RPPs) as a possibility to advance cross-sector integration.

Interviewees felt that focusing on problem-solving (and the cross-sector collaboration that makes it possible) has particular value for data science students who will ideally emerge from their educational programs ready to tackle societal challenges. One interviewee said that they became aware of the Hub through an internship opportunity while in graduate school and now focuses on using technology for social justice. They attributed their career choice at least in part to the real-world experience provided by the internship. They also highlighted a need for a continued focus on connecting research to its various applications, indicating that societal challenges have *“put a spotlight on big data, both positives and negatives of how it can impact our world. We’re starting to see an important need for the Northeast Big Data Hub to be at the center of the conversation.”*

The Impact of the Hub on Stakeholders

Impact measures the perceived added benefit for stakeholders related to collaboration or integration with the Hubs, and the degree to which those who volunteer their time on Hub activities are satisfied with those efforts. We understand this to mean how individual stakeholders rank their own satisfaction based on engagement with the Hub, as well as what overall impact on the data science field can be attributed to the existence of the Hub. Our main findings related to impact include:

- The existence of the Hub has a legitimizing function on the emerging field of data science; and
- Independent of their level of personal satisfaction with the Hub, interviewees presented thoughtful feedback on how the Hub can prioritize its efforts in an attempt to increase its impact.

Various interviewees highlighted the role the Hub plays in legitimizing their work, particularly with institutional colleagues or administrations. They felt that the Hub’s name and reputation carried weight, as did its association with NSF. A few mentioned specific instances where they thought this legitimization occurred, such as using the Hub’s name when creating a new graduate program at the university where they worked, how mentioning the Hub’s support gave credibility to their proposals, or that being affiliated with the Hub had improved their chances of receiving funding or resulted in funded projects. One participant described how being affiliated with the Hub simplifies the task of justifying their work, saying, *“It’s much easier for me. That’s why I usual defer and say, we are working with the National Science Foundation Big Data Hubs.”*

Some participants felt that the Hub legitimizes the data science field as a whole. Data science was described by many interviewees as an emerging field, one that is still taking shape, a process that felt to one like *“building the plane while it’s flying.”* Interviewees spoke about the Northeast Hub and other regional Hubs’ visionary and defining role in this field-building process, expressing opinions such as, *“The Hubs are playing a major role in the emergence of data science as a*

discipline.” Data, and big data in particular, are complicated to work with for many reasons – issues of privacy, compatibility, and ownership all come into play. The Hub, as a legitimizer, can work across these issues and help smoothen the process. One interviewee summed up this thought by saying,

“Having the Hub at the table at those conversations makes it possible. It’s very hard for an individual PI to approach a city or some entity that’s gathering data, so you need a trusted intermediary that makes those conversations possible.”

Impact can also be gauged in terms of how satisfied stakeholders are with their relationship with the Hub. The level of satisfaction across participants varied widely. On the positive side, one interviewee (who has had a close and long-term relationship with the Executive Team) said, *“Out of ten, maybe a ten...just because they’ve been such champions of our work, we wouldn’t be where we are now.”* Another was dissatisfied, citing a lack of clarity about how to engage with the Hub and provide value. Other responses fell across the spectrum, depending on personal experience and individual context.

Virtually all interviewees shared suggestions for the Hub to increase its impact moving forward. While many of these ideas relate to the other four measurement areas (and are discussed elsewhere in this Results chapter), we thought it would be helpful to present them as a single list in this section, as all are intended to increase impact in some way.

Specific ideas suggested by participants include:

- Workshops to educate the public about big data and data science topics;
- Events that bridge academia and industry;
- Greater web presence to spotlight current data science projects and promote broader awareness of data science;
- Regularly scheduled newsletters and virtual meet-ups to keep the field abreast of current research trends, opportunities, and funding possibilities;
- Support for data science standards and defining of common terminology used by the field;
- Access to big data sets by the public and specific communities who can benefit;
- Serving as a clearinghouse or resource archive for data science; and
- Outreach efforts across the Northeast region.

The Hub Is Aligned with the Data Science Community

Representativeness is the extent to which Hub activities align with and support the goals of partner and community projects. We understand this to mean whether and in what ways stakeholders working across the data science community feel supported by the Hub. Our findings related to this measure indicate that:

- Certain stakeholders feel supported by the Hub;
- The Hub is considered an important way to increase the visibility of its members’ activities; and
- Interviewees indicated that having a unified voice for data science in the Northeast Region is valuable for advancing the field.

Perhaps unsurprisingly, those interviewees who had longer and closer relationships with the Hub or the Executive Team felt the most supported. When asked to describe their relationship with the

Hub, one such interviewee, who continues to work with the Hub, acknowledged, *"It's hard to keep it concise, the Hub has been part of my life for the last six years."* The interviewee went on to explain how they and other stakeholders in the Hub came together due to their shared interest in data literacy, and then applied for and received a planning grant. In this and similar cases, our analysis showed a high degree of alignment between Hub activities and the goals of individual stakeholders.

Those without this deeper level of engagement most often thought about the Hub's support in terms of promoting awareness or visibility on an individual and collective basis. They acknowledged working on a wide range of data science projects and initiatives, and saw a role for the Hub to leverage its network to help make those activities more visible.

Collectively, interviewees understood alignment to mean supporting the field to speak together with one voice, especially in conversations between academia and industry. They felt that this was especially important as it relates to the process of making large datasets available for research. One participant commented on this need, indicating there was a general *"lack of appreciation for the complexity of the process leading up to a uniform data set."* They suggested that the Hub could align with those who are already working on this issue, including translational and applied research scientists on the academic side as well as thought leaders from industry. Interviewees thought that doing this would create a feedback loop, increasing the authority of the Hub as spokesperson. *"[If] the Hub represents all institutions in the region it's much more credible, you can get people to show up who would not otherwise shown up to these conversations."*

The Hub Is Positioned for Long-Term Scale

Scale is the extent and depth of change attributable to the Hub's activities, projects, and initiatives. In our analysis, we understand scale to be best represented by the ways the Hub is contributing to accelerating and amplifying the work of data science researchers. Key findings include:

- The Hub's efforts to date have helped advance the field more quickly and decisively than would otherwise have occurred;
- The Hub has been instrumental in promoting broader awareness of certain aspects of data science – such as data ethics – and can provide additional value by amplifying the work of its stakeholders.

Overall, interviewees felt that NSF's investment in the Hub(s) has paid off by helping the field establish a stronger sense of identity. *"I think that data science has gone, during the lifetime of the Big Data Hubs, from a 'What are you doing?' to a 'Oh yes, this is one of the most important things we could be doing.'"* Several stakeholders acknowledged that the field is maturing quickly, indicated in part by the growing number of programs in universities, and attributed some of this acceleration to the work of the Hub.

They acknowledged that many projects and initiatives could have happened without the Hub's involvement, but that the Hub's efforts have helped speed up some of the work. As one participant said,

"It's hard to pinpoint something and say that would never have happened...but there are lots of places where things have moved along

faster and come to fruition in ways that they would not otherwise have if the Hub were not in place.”

Another participant emphasized the Hub’s role as a catalyst for certain types of research projects noting that, *“The Big Data Hubs have been an astoundingly good and successful way of getting science to turn on a dime, so to speak.”*

We also found evidence suggesting that Hub efforts have helped some researchers think differently about the data science field, and they are sharing that knowledge with others in the community. Following a Northeast Hub in-person meeting, a university-based researcher commented that the meeting *“actually put me in a whole new path of research I hadn’t been in, which is ethics and data science, which has proven to be very important.”* The researcher went on to describe a project where they had doctoral students try to explain the ethics of their work. This exercise led them to realize that the students lacked formal training or exposure to the ethics component of their work. The interviewee mentioned the *“scalable value”* of taking what they learned at the Hub event, and working it more systematically into university practice.

A small number of participants noted that the Hub helps to disseminate and / or publicize their work, and that without the Hub, the data science community would lose a space to highlight projects and research that haven’t been published or shared broadly. Interviewees felt that the Hub’s capacity for amplifying the efforts of stakeholders would result in scalable impact. One interviewee noted that this could be done through co-branding efforts, and taking advantage of other established campaigns that already have traction (such as Cybersecurity Awareness Month). Two participants who are currently involved with research networks noted the potential for these and other networks to amplify or promote further awareness of the Hub and its activities. They described it as a *“two-way street”* where they can provide the Hub with input from the data science communities they work with, while promoting the Hub’s messaging across their networks.

Discussion & Recommendations

Overall, responses from both the Community and Project participants interviewed for the evaluation suggest that the Hub is making good progress towards its stated goals. These data suggest that Hub occupies a central role, one that positions it to engage stakeholders from various sectors of the data science field. And it is able to do so in a way that brokers deeper and richer collaborations than any individual institution would be able to do. In large part due to its role as a convening entity, respondents felt that the Hub would be well suited to address some current gaps in the field primarily around gathering and sharing large datasets. They also observed that there a few current areas of activity where the Hub could do more to maximize its efforts.

Build collaborations to address real-world challenges through translational data science approaches

There is a significant amount of evidence from the responses in support of the Hub’s role in enabling collaboration, often between individuals who might not otherwise have connected. Participants mentioned conversations during Hub-organized events or named specific members of the Hub’s Executive Team as crucial to facilitating interactions. This web of connections made through the Hub has resulted in funded projects as well as fostered collaboration across

disciplines, institutions, and sectors. One important benefit of brokering relationships between academia and other sectors like industry, government, and nonprofits is the opportunity to leverage these partnerships to address real-life societal challenges. Several participants mentioned the possibility of using research-practice partnerships as a way to accomplish this.

Given the importance of its role as a connector, several participants interviewed for this evaluation wanted the Hub to have a much stronger presence across the Northeast region, voicing a desire for clarity around its efforts to engage stakeholders. This suggests that there is room for the Hub to be a bit more strategic in its communication approach and perhaps leverage a broader range of outreach mechanisms to inform the community about planned events and activities. We offer some suggestions for this purpose in the recommendations section below.

Foster innovation and scale endeavors that reflect regional interests and align with national priorities related to data science

While we cannot really speak to regional interests or national priorities as this was not the focus of the evaluation, we can speak to innovation and scale. In general, interviewees felt that the Hubs as a whole have contributed, at least in part, to the rapid maturation of the data science field. Although the exact particulars of how the Hubs have accelerated the field are a little unclear, some participants were able to mention specific instances where their affiliation was of benefit. For example, a participant mentioned that their affiliation with the Hub was supportive in creating a data science program at their university. Several participants also described the Hub as a catalyst for research projects. Related to this, various interviewees highlighted the Hub's role in legitimizing their work to institutions and funders.

A few participants noted the role that the Northeast Hub plays in spotlighting projects and research that had not been published or disseminated broadly. In addition to increasing the visibility of data science projects, this particular role has implications for the Hub's goal to foster collaborative innovation moving forward by providing an alternate – yet highly visible – space for sharing the work of its stakeholders, beyond traditional academic outlets. Moreover, spotlighting projects in this way can lead to new funding opportunities or help collaborations target real-world problems.

Support and promote representative community engagement / impact across all Hub activities

Although the levels of awareness and involvement varied, evidence from this first round of interviews suggests that the Northeast Hub is working to engage members of the data science community across the board. Several participants specifically highlighted the Hub's role in building bridges between data scientists and practitioners within academia as well as between academia and industry. Most felt that the Hub has made progress toward creating a vital and interconnected network that would be lost if the entity were no longer there. And many recognized that there were ample opportunities to bring in stakeholders from other sectors, including more industry players as well as people from non-profits, government, and even the general public. Partnering with libraries who already offer programs for the public was one of the ways participants felt the Hub could engage more deeply with non-academic audiences.

Increase data science capacity and talent, emphasizing underserved communities.

The general consensus from our analysis of the interview responses was that the Hub has an important part to play in increasing data science capacity, access, and opportunity for the broader community of potential stakeholders. Participants recognized the Hub's past efforts to engage the data science community through various activities including in-person and virtual events. This is work that many hope will continue and expand. Some of the concrete suggestions that came out of the analysis include promoting specific opportunities that are designed to increase participation from students, and to better engage with communities that have been traditionally underrepresented in fields related to data science. One concrete way that the Hub could increase students' participation is by connecting them to more internship opportunities. These internships could focus on academic applications of data science or on applications of data science methods to important societal challenges. In terms of general accessibility, one suggestion was to co-sponsor workshops and events in commonly spoken languages like Spanish with partners who are already engaged in that work.

As noted above, participants felt that the Hub could refine its communication strategy around activities and opportunities to encourage greater involvement from all facets of the community, and to better connect participants in regions of the Northeast outside of New York City. Besides awareness, another frequently mentioned barrier to engagement was time. In the recommendations section, we suggest ways that the Hub can navigate concerns about time constraints.

Recommendations

Based upon the interviews, we provide the following recommendations for expanding the impact of the Northeast Hub, understanding that some of these ideas may be best pursued jointly with other regional Hubs. We recognize that some of these recommendations are already occurring to various degrees. However, interviews made it clear that either the community lacks awareness (indicating a need for refining the Hub's communication strategy) or the frequency / depth of these activities should be increased. Wherever possible, acknowledging funding and staffing limits at the Executive team level, we have sought to identify ways that the Hub can leverage its impact through the work of its partners.

- **Community education events:** Co-sponsor / promote workshops, online classes and / or events to educate the public about big data and data science topics, prioritizing venues that are widely accessed by non-academic audiences whenever possible;
- **Bridge academia with real-world applications:** Co-sponsor / promote ways – events, discussion forums, etc. – that allow for academia to find applications across other sectors, such as nonprofit, industry, and government;
- **Web presence:** Utilize the new website and other social media tools to have a stronger web presence to promote broader awareness of data science, making use of existing networks with listservs to increase the reach of communications;
- **Sharing of new developments in the field:** Whether through regularly scheduled newsletters, the website or other platforms, become the distributor of current research trends, opportunities, and funding possibilities;
- **Flexible Scheduling:** Working with partners to vary the times when planned events occur or offering multiple versions of the same events may give more members opportunities to

participate that fit with their schedules. Another option might be, where possible, to record events and make these publicly available for participants to view in their own time.

- **Promote diversity, equity and access:** Reach out to stakeholders who already have a focus on inclusivity and promote their efforts. Use internships and convenings as an opportunity to bring younger and newer voices into the field.
- **Contact points:** The Hub has access to a wealth of talent and expertise that can be utilized to a much greater degree. Create clear mechanisms for interested individuals to get in touch with the Hub and guidelines for when and why to reach out, along with specific ways they can help.
- **Volunteer-led working groups:** We concur with the suggestion from participants to clearly establish and promote working groups around different aspects of data science, led by volunteers from the community who are experts in those domains. This creates another opportunity for like-minded individuals to connect and collaborate on topics and issues of interest to them, and can be an additional way to engage new members.

Interviewees emphasized the importance of several larger undertakings that they saw as priorities for advancing the field. While we believe these to be outside the scope and capacity of the Northeast Hub, we note them as important long-term opportunities for strengthening the field. We encourage the Hub to consider its role in addressing these greater challenges.

- **Clearinghouse:** Become a clearinghouse or resource archive for data science;
- **Standards Designer:** Design and promote standards for data science and coordinator for common terminology used by the field; and
- **Data Repository:** Become a central place for the sharing of or access to big data sets.



Response to Evaluation: Round One

In October 2020, the Northeast Hub Executive Team provided a response to Knology’s mid-point evaluation report (Ardalan, et al., 2020), which was included as an appendix to that publication. The black text below reflects the content of that response. In February 2021, the Executive Team provided Knology with an update on the Hub’s activities to date. Those updates are incorporated into the Hub’s original response, where additions are indicated with blue text below.

Letter to Knology in response to initial evaluation report of the Northeast Big Data Innovation Hub

The Northeast Big Data Innovation Hub extends our gratitude to the team at Knology for their work in conducting the first phase of their evaluation of NSF award 1916585. We look forward to keeping the team up-to-date on current and upcoming activities at the Hub, and to the insights they will provide in the second phase of the evaluation, slated for spring 2021. The below summarizes our response to the findings of the evaluation’s first phase, including specific actions with timelines being taken by the Hub’s Leadership team.

The mission of the Northeast Big Data Innovation Hub is to build and strengthen partnerships across industry, academia, nonprofits, and government to address societal and scientific challenges, spur economic development, and accelerate innovation in the national big data ecosystem. Our collaborations with this community are critical to this work. Consequently, we were delighted to hear confirmation in this report that our efforts provide real value to our community – including effectively facilitating cross-sector networking and collaboration, organizing activities that align well with the priorities of our stakeholders, and supporting our community by amplifying their work.

The Executive Team has discussed Knology’s list of recommendations (see Appendix) for enhancing our existing impact, and how we build them into our strategy and operations. Our guiding principles for this work include:

- Be **proactive** in engaging with our community, listening to them, and inviting them to get involved. This includes proactive leveraging of other channels to our constituency, including partnering with Regional Research and Education Networks, and developing the COVID Info Commons Community.
- Be **accessible and inclusive**, reaching individuals and institutions of all kinds, from all parts of our region.
- Be **focused**. Our June 2020 strategic plan identified four focus areas, consolidating from a more widespread set of 8 after research and discussion with community stakeholders. By focusing our efforts, we grow our strengths and proved easier access points for engagement.

Specific action in alignment with these principles that we have taken, are taking, or plan to take within the next quarter, include:

We will continue to develop community programs and organize events that serve as a bridge between academia and other sectors. These include the Northeast Student Data Corp Founding Committee, designed to bring together leaders from all sectors and students of all backgrounds

to develop a program teaching data science fundamentals to underserved populations. The initial cohort for the committee convened for the first time in October 2020.

As of February 2021, the Hub has taken the following actions:

- Launched the Northeast Student Data Corps - [Data Science Virtual Career Panels](#), and communicated broadly to high school and community STEM programs, 150 academic institutions, directly through outreach emails to community colleges, career services, data science institutes, vice provosts of research, regional education networks, libraries. Registration is at 500 individuals as of Feb. 4, 2021 for the inaugural event scheduled for Feb 19, 2021.
- [Presented at Iona College](#), virtually – the school currently does not have a data science program.
- Co-sponsored the Boston Areas Research Initiative [2020 BARI Conference](#), Reimagined which was held as a series from April through July 2020.

Beyond our existing community events, we will reach broader and non-academic audiences by:

- Participating in existing conferences and events popular among the community as well as reaching communities we may not yet know. As a recent example, the Hub submitted a birds of a feather (BOF) proposal on diversity, equity and inclusion to the Academic Data Science Alliance for October 2020, and offered it as a base for a larger BOF. Leading on this has resulted in combining the Chan Zuckerberg team in the BOF, and Fred Hutchinson Cancer Research Center.
- Organizing new community events. These include monthly COVID information Commons (CIC) Community webinars, beginning with the CIC launch in July 2020 and continuing with monthly COVID-19 lightning talk webinars with invited COVID RAPID PIs from around the country, which kicked off in September 2020 and are slated to run through Q1 2021. We plan a CIC student paper competition by early 2021 to engage more students in the COVID-19 research community.
- Co-branding existing and new events with collaborator organizations, extending our reach. These include a Connected Healthcare Cybersecurity workshop cosponsored with IEEE, for which an organizing committee was first convened in October 2020. The event is expected to be held in Q1 2021. Other co-branded events include leveraging our partnerships with regional RENs (Research and Education Networks) such as a webinar with KINBER for Pennsylvania schools and libraries held in September 2020, and a future event we are planning with NYSERNET for education and research communities in New York state.

As of February 2021, the Hub has taken the following actions:

- Florence Hudson, the Hub ED, co-led a BOF [on Diversity, Equity and Inclusion – The Path Forward](#) at the ADSA Annual Meeting on Oct 16, 2020, along with the Fred Hutchinson Cancer Research Center and the Chan-Zuckerberg team.
- The Hub supported the [Northeast Student Data Corps](#) founding committee [inaugural Data Science Career Panel](#) which included data science professionals from Esri in Industry, New York City - Mayor's Office for Data Analytics, U.S. White House Fellow, researchers in bioinformatics and real time earthquake prediction/risk mitigation
- The Hub hired a Columbia University grad student in data science to develop use cases to add to the NSDC and [Data Science Resource Repository](#).

- The Hub held 7 CIC webinars from July 2020 through February 2021, hosting 500 attendees and 44 PI lightning talks, bringing together researchers, students, government, academia, industry and nonprofits to enable researcher collaboration for COVID.
- The Hub has is partnering with [*IEEE Standards Association to co-sponsor a Connected Healthcare Cybersecurity virtual workshop series*](#). The series has been launched, with its first workshop on Feb 24, 2021 followed by 4 more workshops scheduled through November 2021. It will develop standards recommendations and roadmaps for connected healthcare device and data security, privacy, ethics and interoperability, working with academia, industry, government, nonprofit, researchers and students.
- Co-branded events with the regional RENs (Research and Education Networks) such as a [*webinar with KINBER*](#) for Pennsylvania schools and libraries held in September 2020, and on a [*New Jersey EDGE conference*](#) in January 2021.

We will continue to leverage our web presence and share new developments in the field by:

- Creating more content highlighting Northeast data science projects, outcomes, and opportunities, from community voices – e.g., success stories from our community and interviews with COVID-19 researchers, which launched during the summer of 2020.
- Incorporating topics highlighting new developments in the field, e.g., insights from CIC COVID-19 researchers, and new IEEE cybersecurity standards efforts.
- Expanding existing multi-channel communications – web, Twitter, LinkedIn, YouTube, Northeast Hub and COVID Info Commons community newsletters, and program-specific channels such as Slack – on an enhanced production schedule deployed in August 2020.
- Meeting quarterly to review engagement analytics to inform communications strategy, with the first meeting slated for October 2020.

As of February 2021, the Hub has taken the following actions:

- <http://nebigdatahub.org/> was relaunched in June 2020 following the release of the Hub's new strategic plan, clearly communicating the Hubs' new and existing programs and activities in alignment with this plan.
- Uses Handshake to advertise events such as the Northeast Student Data Corps. activities, data science career panel, and student employment/leadership opportunities.
- Partnered with regional education networks such as KINBER in Pennsylvania and Edge in New Jersey to be channels for Hub communications and activities.
- Increased usage of Twitter including tagging other networks such as NSBE (National Society of Black Engineers), SHPE (Society of Hispanic Professional Engineers), SWE (Society of Women Engineers), AISES (American Indian Science and Engineering Society).

We are operationalizing our mission statement on diversity, equity and inclusion, created in the summer of 2020, through intentional outreach to and collaboration with underserved communities on an ongoing basis throughout all aspects of our programming. As a recent example, we reached out to 84 new institutions across the Northeast to invite participation in our 2020 seed fund, of which 32 were HBCUs, primarily black institutions, and Hispanic-serving institutions. The outreach list was built to encourage inclusion from a demographic perspective, and also geographically (reaching every state in the Northeast) and institutionally (reaching universities and two- and four-year colleges). [We continue to add to the outreach list, with new institutions as well as new contacts at those institutions, for instance adding career services contacts for the Northeast Student Data Corps - Data Science Career Panel.](#)

We will develop programs that support underserved communities, such as the Northeast Student Data Corps, in partnership with collaborators from these communities.

As of February 2021, the Hub has taken additional actions in this regard:

- Created a [*Diversity, Equity and Inclusion mission statement*](#) on the Hub "About" page.
- Created an outreach list of 150+ academic organizations, including 40+ Minority Serving Institutions (MSIs), with details on data science programs, vice provosts for research, and careers services offices for each institution. This outreach list is used for new program invitations, including the Seed Fund, CIC Student Paper Challenge, and Data Science Career Panels, resulting in increased engagement.
- The Hub community now includes 93 Minority Serving Institutions who receive our newsletters, outreach lists, and are invited to participate in our programs and events, including Hispanic Serving Institutions (HSI), Primarily Black Institutions (PBI), Asian American and Native American Pacific Islander - Serving Institutions (AANAPISI), Historically Black Colleges and Universities (HBCU), Alaska Native and Native Hawaiian - Serving Institutions (AANH), Native American - Serving Nontribal Institutions (NASNTI), Tribal Colleges and Universities (TCU).
- Increased outreach to high school STEM programs in minority communities to invite them to Hub programs, including the Northeast Student Data Corps and Data Science Career Panel.

As an ongoing commitment to our community, we leverage flexible scheduling by offering multiple time slots and deadlines for Hub events and programs, as staffing and community interest allows, and schedule events to maximize participation from the Hub community. We will continue to leverage live streaming and recording to share our events more broadly via out multiple social media channels including the website, YouTube and Twitter.

As of February 2021, the Hub:

- Supports the COVID Information Commons (CIC) website which includes 20 [*research funding opportunities*](#) globally from government, industry and not for profits. The COVID Information Commons webinars include webinars have included 40 [*research lightning talks*](#) to share current research trends in addressing the impact of COVID-19.
- Sends out the Hub newsletter which communicates about NSF funding opportunities.
- Sends out the Monthly Northeast Hub and COVID Information Commons newsletters which include news, events, and opportunities.

In addition, the Hub mitigates scheduling conflicts for members through:

- Events and webinars scheduled at varying times to maximize participation.
- Recording all webinars and posting them on [*CIC videos*](#) and Northeast Hub website.
- For those who can attend, webinars are live streamed on YouTube. They are also posted on [*Northeast Hub YouTube channel*](#).

We have integrated organization-level touch points for contact and collaboration throughout all aspects of our programming – as an example, see our form to attract new collaborators at nebigdatahub.org/collaborators. We will continue to leverage program-specific touch points, such as community Slack channels, topical webinars, and topical working groups with open calls for community participation. These include ways of contacting the Hub, but also ways for community members to contact each other.

As of February 2021, the Hub has taken the following actions:

- Hub web pages and monthly newsletters all have a mechanism to contact the Hub.
- The “About” webpage with the four focus areas and projects has an [*“Add your project here!”*](#) button which provides a form to add their project to the Hub community.

To further our commitment to community participation, we disseminate open calls for community participation in volunteer-led working groups that further the Hub’s mission. Current and future volunteer-led working groups include the Hub’s Steering Committee, Advisory Board, Seed Fund Steering Committee, Northeast Student Data Corps Founding Committee, Connected Healthcare Cybersecurity Working Group with collaboration with IEEE on responsible data science for clinical IoT and cybersecurity, and Urban to Rural Communities Working Group, all of which are already active. We are building such opportunities for community participation in all future community projects.

As of February 2021, the Hub has taken the following actions:

- Created new Hub working groups such as [*Urban to Rural Working Group*](#), which includes a Water Management subgroup and Food Security subgroup; as well as a Connected Healthcare Cybersecurity Workshop group in collaboration with IEEE.
- Connected healthcare cybersecurity workshop group in collaboration with IEEE.
- Is leading collaboration with the American Geophysical Union for all four Hubs to address water management challenges with increased data sharing, supporting the water subgroup of the Urban to Rural Communities Working Group. The water subgroup includes academia, nonprofit, and agency related efforts such as the Department of Energy’s NAWI program, the [*National Alliance for Water Innovation*](#).

Given the additional recommendations highlighted by the findings, we will explore roles for the Hub as a clearinghouse, standards designer, and/or data connection repository, leveraging existing programs discussed above. Building off our strength in community engagement, we will explore opportunities – particularly through volunteer-led working groups, web presence, and community events – to share resources and data sets and promote the design of standards and common terminology. We will publish Knology’s report, including our response, on a dedicated web page, outlining our commitment to our community, our ongoing actions to scale our value to the community, a “fact sheet” on community engagement metrics and a feedback mechanism for anonymous community input. We expect this page will be launched by December 2020.

As of February 2021, the Hub has taken the following actions:

- Created a [*Data Science Resource Repository*](#) which includes links to open datasets.
- Partnered with IEEE Standards Association (Institute of Electrical and Electronic Engineers) and the *IEEE/UL (Underwriters Laboratories) P2933 standards working group* on clinical Internet of Things data and device interoperability with TIPPSS (Trust, Identity, Privacy, Protection, Safety, Security) to co-sponsor the [*2021 Connected Healthcare Cybersecurity Workshop Series*](#).
- Supported the [*COVID Information Commons*](#) which includes links to 50 open datasets.
- Posted the Knology evaluation report on the Hub “About” page in “Community Engagement and Progress” <https://nebigdatahub.org/northeast-big-data-innovation-hub-progress/>.

We are grateful for the many opportunities we have to serve our community, and consider our work in doing so to be ever-evolving, with our goal to always learn and improve. We thank

Knology for their role as evaluator on this project, and look forward to future insights from them and, as always, from our community themselves.

Jeannette Wing, Principal Investigator

Florence Hudson, Executive Director

Katie Naum, Operations Manager



Evaluation: Round Two

Methods

Participants

Round two of the evaluation shifted away from the original intent to interview the same group of stakeholders a year after the first round of interviews – in part to reflect the vision, focus, and activities of a new Executive Director, Florence Hudson, who assumed the role in May 2020. Instead, Knology interviewed data science researchers and practitioners who are more actively involved with the Northeast Hub’s current projects, some of which continue from earlier days of the initiative and some of which are new initiatives not included in the original grant proposal. This approach allowed evaluators to gain deeper insight into timely collaborations – including a robust response to COVID-19 – rather than re-assess historical relationships that may no longer be as relevant to the Hub’s work. We conducted a total of 12 interviews in Round 2, four of whom had been repeat interviewees from Round 1.

Instrument

The Round 2 interview questions were similar to those used in Round 1, but updated to reflect interviewees’ participation in more recent Hub activities and initiatives. In cases where Round 2 interviewees had also been interviewed in Round 1, researchers returned to their original responses and asked questions that followed up on what interviewees said in their first interview.

Analysis

A researcher conducted all interviews and performed the thematic qualitative analysis. Two additional researchers were involved in conducting and synthesizing conversations with Hub leadership to provide context for the evaluation findings, most notably regarding the change in executive leadership and the launching of new initiatives following the first round of evaluation. The research team organized the results to reflect the five central measures for the project, as defined on page 3 in this report.

Results

Recent Progress: Reach & Engagement

Reach / Engagement measures the number and variety of communities affected by Hub activities and the number and variety of partnerships the Hub works with. In the first round of evaluation, interviewees said that they benefitted from past events and applauded the Hub’s outreach efforts, yet were not always clear about opportunities for future engagement. The second round of evaluation builds considerably upon findings from the first round related to this measure:

- The Hub's more streamlined and targeted outreach efforts have resulted in deeper and more diverse engagement with the community;
- Members appreciated the improvements to Hub-affiliated websites and found them to be useful resources; and
- There are clear channels of communication between the Hub and members of the data science community.

The Hub recently updated its website to improve its functionality and provide more targeted communications to stakeholders. Interviewees found the recently updated Northeast Hub website (<https://nebigdatahub.org/>) to be a useful resource. Various statements supported this finding, such as *"I feel like they keep adding features to the website and that's really helpful."* Interviewees also said they increasingly trust the website to provide relevant and timely information.

Respondents expressed similar sentiments about the new COVID Information Commons website (<https://covidinfocommons.net>), a resource that explores NSF-funded research related to the pandemic. One faculty member remarked that *"students know they can find out more about who's doing what with COVID and rapid grants from NSF."* Interviewees familiar with both websites noted that they provide valuable information and are clear and intuitive to navigate. There is also evidence that the interviewees recognized that the website was more than just a source of information. They also saw it as a way to connect with the Hub staff who they viewed as approachable, well-informed, and responsive. One interviewee said that they would feel comfortable emailing Hub staff and asking for additional clarification about information on the site if necessary. Interviewees also liked the idea of highlighting success stories or community spotlights through the Hub website and newsletters. They noted that these stories would position the Hub in a positive way and provide valuable exposure for the data science community as a whole.

Most interviewees thought that the Northeast Hub-affiliated websites struck an appropriate balance between providing pertinent information and overwhelming users with too much detail. In addition, they noted that the sites avoided another potential pitfall – that of collating information without making it useful. One person said, *"You gather and you gather and you gather all this information, but if you don't synthesize and you don't use it, it's of no value!"* The perceived value of the COVID Information Commons site in particular was this synthesizing function – what one interviewee described as a critical role for the Hub more broadly as well. Well-structured websites were one component of the Hub's strategy for connecting with the community. One interviewee remarked that the *"website is [a] necessary but not sufficient condition for communication"* and that other engagement strategies are necessary in parallel. One such strategy that the Hub currently uses are "lightning talks" which are given by members of the community. Beyond giving initial exposure to data science researchers working in diverse domains, interviewees saw the lightning talks as a useful foundation for continued collaboration as well as greater reach and engagement. Several interviewees in this round said they had previously given lightning talks.

Another strategy for reaching and engaging with a broad range of stakeholders that the Northeast Hub has adopted are the seed grants awards. One interviewee commented on the Hub's intentional efforts to engage with the community in general and with minority-serving institutions and community colleges in particular through outlets such as a targeted email outreach to individuals identified as working in data science in the region. An interviewee with

knowledge of the seed fund process felt that it typified the Hub's commitment to diversity and inclusivity, as well as support for those underrepresented in data science fields.

Lastly, student engagement was an area of recent improvement for the Hub that several interviewees mentioned. They described the Northeast Hub as a place that actively invites and encourages student engagement and participation, through opportunities such as the Northeast Student Data Corps and a COVID Information Commons student paper challenge – an initiative that was designed by students.

Overall, interviewees spoke very positively about the Hub's efforts on reach and engagement. They noted that there seemed to be more opportunities available now for members of the community to participate, that those opportunities were clearly communicated, and that engagement efforts felt more *"strategic"* and they had the sense that *"someone is listening."* Beyond what the Hub has already done, ideas to sustain this broad reach indicated a preference for more visual content – such as short video updates, asynchronous events aimed at promoting the Hub's activities or where Hub leadership talk about the vision. Some respondents suggested engaging in exchanges using popular formats, like a PI "Ask Me Anything" on Reddit, known colloquially as an AMA.

Recent Progress: Integration & Interconnection

Integration or **Interconnectedness** measures the extent to which partnerships amplify and cross-pollinate Hub-related activities to create economies of scale. It focuses on how the Hub supports cross-sector collaboration, reduces disciplinary silos, and connects academic research with real-world application. The second round of evaluation builds on findings related to integration or interconnectedness from the first round of evaluation. In the first round, interviewees highlighted the efforts of the Executive Team to facilitate collaboration within academia as well as between academia and other sectors. Key takeaways from the second round focused on renewed efforts in this area and highlighted recent activities of particular benefit to stakeholders:

- The Executive Team has stepped up efforts in the area of community building, yielding tangible outcomes;
- Regular lightning talks have amplified the community's research and helped strengthen connection; and
- Through the Hub's activities, members have expanded their professional networks and been exposed to new opportunities for collaboration.

Promoting interconnectedness is seen as a particular strength of the Hub. While there has been demonstrable evidence of successful networking throughout the Hub's existence, heightened success in this area was directly attributed to the recent change in executive leadership. One person commented, *"They're doing a darn good job of linking people together, and that's the best use of data."* About working with the current Executive Director Florence Hudson in particular, one respondent reflected, *"It's been a very positive, very pleasant and illuminating kind of relationship."*

Many interviewees spoke about new connections that they made through the Hub and its activities. One interviewee stated, *"One of the hallmarks of the Northeast Big Data Hub is the collaboration aspect."* They described this culture of stakeholders working together as part of a larger entity as fundamental to the Hub's identity. Respondents noted that these new connections

and collaborations expose researchers to a range of potential applications for their work. One interviewee found that the partnership network built up around the Northeast Hub has resulted in *“a regional center of excellence”* for data science research.

The Hub’s efforts to foster integration and interconnectedness has resulted in distinct outputs: workshop presentations, proposals, and various other collaborative efforts. As an example of the kinds of long-term connections that are being made, two academics met at a Hub event. Two years later, one invited the other to be part of a workshop. In another case, two individuals met through the Hub and realized that they were working on a similar topic – support for public libraries and data science. They have now built a community of people interested in this topic area within the Hub, with an eye toward continued collaboration.

The new practice of hosting regularly scheduled lightning talks illustrates the amplification effect of interconnectedness. One interviewee stated, *“I think that the lightning talks have been really very beneficial for people to engage with one another, as well as I think that’s where you get connection with people you don’t know otherwise.”* For example, one interviewee who has given a lightning talk noted that it was an opportunity to get input from the community, learn about other areas of research, and to meet a range of potential partners. Another interviewee noted that their talk resulted in new collaboration for a grant proposal. These talks are also creating new opportunities for students to access and analyze data. One interviewee who had given a lightning talk organized a similar event for students that had participation from many schools.

Supporting the growing network requires both coordination and clear communication. For some interviewees, the new strategic plan provides a roadmap as well as focused communications about the Hub’s vision, mission, and strategy. The strategic plan also provides a way for Hub leadership to talk about its work with stakeholders and partners. Interviewees noted that Hub staff routinely pointed partners to the new strategic plan in conversation as a reference for understanding the Hub and how it supports partners. Some noted that the updated strategic plan had helped them understand the Hub’s vision and mission. One interviewee noted:

“Now that the new strategic plan is very clear and is being actively promoted to members of the community, it is very clear to me what the new direction of the Hub is and the leadership and vision, and my role in it and my organization’s role in it.”

When asked what would happen if the Hub did not act as a coordinating entity, one interviewee said:

“To have that regional impetus for collaboration around data science, big data, and data education. It wouldn’t be happening in this sort of coordinated way... We’re very fortunate to have the Hub in the Northeast, that we have this convener and this catalyst. So if the Hub went away that’s what would be lost, plus all the energy that [the Hub leadership] brings would be missed greatly!”

Not all interviewees were aware of the four Hub regional structure, but those who were mentioned that they had recently seen a change in the way Hubs interacted with each other, with implications for progress on this measure related to “economies of scale.” To an extent, interviewees saw the increased successes of networking, communicating, and coordinating among the Hubs as resulting directly from the efforts of the Northeast Hub’s Executive Director.

Some interviewees also noted that the community building efforts by the current Hub leadership provided an antidote to some of the feelings of being disconnected from others during the COVID-19 pandemic.

A critical aspect of integration and interconnectedness measures that emerged in conversations with interviewees is interdisciplinarity. Several interviewees described this as a priority area for the Northeast Hub's efforts, including thinking about the range of application areas and communities that would be involved. They were clear about the need to bring disciplines together to work on solutions, while acknowledging some of the difficulties of interdisciplinary collaboration. As a discipline that crosses multiple academic areas, data science requires a level of cross pollination across different fields that universities are often not able to adequately address. One interviewee acknowledged that this would not naturally happen, if not for the Hub:

"Facilitating these kinds of meetings, saying, 'hey you're working on water, we're working on water. Why don't we put you in a room and you talk to one another?' That's something that's really important and it would certainly be lost."

Recent Progress: Impact

Impact measures the perceived added benefit for stakeholders related to collaboration or integration with the Hubs, and the degree to which those who volunteer their time on Hub activities are satisfied with those efforts. In round one, participants recognized the legitimizing function of the Hub on the data science field as well as provided thoughtful feedback for ways the Hub could prioritize efforts to increase impact. Progress reported in round two of the evaluation was as follows:

- Interviewees indicated higher levels of satisfaction with the Hub and its efforts than they did in the first round;
- Interviewees noted a few specific areas of impact such as the COVID Information Commons and seed grants; and
- Interviewees noted a need for continued NSF support in the immediate future as the Hub's impact continues to grow.

Interviewees noted very high levels of satisfaction with the Hub and their involvement in Hub activities as evidenced by the subset of comments:

"It's been really great!"

"To the highest level!"

"I feel very satisfied! Definitely."

Those interviewed for a second time were more satisfied than they had indicated in the first round of interviews. When asked to rate their level of satisfaction, interviewees elaborated on the reason for their rating with comments such as:

"They're very responsive, always open to conversation and discussions, brainstorming about new ideas, open to expanding into whatever areas might be useful for the researchers involved."

"When I see what they did with their support and the seed funding and getting early career young folks involved, I just think that's fantastic!"

“Whenever I’ve had to reach out somebody gets back to me in a timely manner.”

(Speaking about the change in executive leadership) *“Now the Hub seems like something that might yield what I need.”*

Other interviewees were satisfied due to perceived benefits for students participating in the Hub (such as the opportunity to be a co-PI on a seed fund grant), or the exposure from being involved with Hub leadership at prestigious conferences. Some reasons for satisfaction had to do with more operational concerns – perceiving Hub staff as being excellent communicators who are accessible and responsive. Interviewees also expressed satisfaction with seeing the Hub research projects have real-world impact. A frequently cited example of research to practice to a societal challenge that demonstrated how the Hub has helped enable data availability was the COVID Information Commons initiative. Respondents described the resource as having a *“clear effect on the pandemic and understanding COVID.”*

In a relatively short amount of time, seed fund awards have shown demonstrable impact across the Hub’s network. Seed grants help establish a funding track record so that future grant applications are more likely to be successful. Evidence from the evaluation suggested that these grants are particularly essential for students with financial hardship whose participation in data science activities may be otherwise limited – even a small amount of money can provide the extra incentive needed. One interviewee commented that while \$25,000 may seem insignificant to a university that routinely receives millions in grants, it can kickstart a young person’s academic or professional career. With that money, faculty can support students to do data collection and analysis, buy software that will advance the efforts of a team, or use the grant as leverage for a future award. Another interviewee noted that greater community engagement with the Hub is contingent on being able to provide a *“financial carrot”* as an incentive.

To scale the seed grants award program, one interviewee thought that the Hub needed to provide greater clarity and responsiveness regarding budgets, expressing a desire for a budget template. Another sought better alignment between the timing of awards and the academic calendar, commenting on the accelerated timeline once funds were disbursed, a challenge exacerbated by scheduling conflicts with the academic calendar. They suggested changing the timeline to be more compatible with the academic year so as to not limit the usefulness of the funding.

For impact to continue and increase, interviewees were clear about the need for continued NSF support, at least in the immediate short-term. They described the Hubs’ role in the data science community as both unique and critical, and they worried that a detrimental gap would exist if the Hub were to disappear. The following quotes from interviewees best speak to the perceived impact of the Hub:

“The Hub is really something unique...it cannot be easily replaced by something else.”

“When you want to look at who are the nation’s thought leaders in data science you would naturally look at the leadership of the Hubs, the director and their advisory boards.”

“When I look at Hubs – and I look at them nationally – they are really the pinnacle for NSF. ... The directors typically will advise on other

matters related to data science nationally. That's what I meant by pinnacle. If I'm looking for leadership across the nation, it will occur in these Hubs."

Some interviewees also claimed that the US needs to provide focused support for data science research in order for the country to remain competitive globally across different sectors that increasingly rely on big data as part of their business model.

Recent Progress: Representativeness

Representativeness measures the extent to which Hub activities align with and support the goals of partner and community projects. In the first round of evaluation, interviewees described the Hub as providing a unified voice for data science in the Northeast and also helping to boost the visibility of its members activities. Recent progress captured in round two of the evaluation reflect the change in the Hub's Executive Director and new strategic plan, with results indicating that:

- Stakeholders could more clearly articulate how the Hub supports their work; and
- There is a greater sense of alignment with the Hub's mission and vision among members of the community.

Overall, interviewees described a sense of mission alignment with the Hub as it exists currently. They noted that Hub staff are generally enthusiastic, highly responsive, and demonstrably supportive. Some interviewees specifically called out the efforts of the Hub's Executive Director. One said simply, *"I've never received a 'no' from Florence!"*, commenting on what they found to be an exceptional willingness to listen and support partners' efforts. Another said, *"We think that the Hub and Hub activities are important to our community...[the Hub] seems to be very aligned with our mission and activities, and I don't see that changing."*

Evidence suggests that the Hub's efforts add credibility to institutional pursuits for those working in the academic sector. In one case, an academic that was trying to build a data science program found that association with the Northeast Hub – and the top-caliber network of its leadership – made that process smoother. Others substantiated this idea, noting that the Hub is a place to unite efforts and find like-minded partners. One interviewee remarked,

"The region will do a lot better if we're all working with each other, if we're all supporting each other. For my institution, it's really been recognized that, wow, you've been able to come in and work closely with the Northeast Big Data Hub. That gives us credibility. It gets us a seat at the table with some great colleagues around the region, and that's really been valued."

Interviewees expressed appreciation at having learned about other valuable and relevant grant opportunities through the Hub. They specifically highlighted the seed fund awards, noting that these directly support partners' goals, as do communications regarding additional grant opportunities and invitations to join proposal teams. In one case, receiving a seed grant meant that the awardee also received the institutional backing needed to proceed with a large project. As they said, *"If I didn't have an available grant, they would have told me to wait until I got a grant] ...That's why its name is a 'seed' grant, so you can show you can get a grant."* The same individual expressed appreciation at having learned about other valuable and relevant grant opportunities through the Hub.

The Hub promotes members' work through various communication strategies and channels, which has led to tangible outcomes for some partners. As one interviewee said,

"The Hub has been great about getting us press and writing stuff up, that's been very helpful...One of the things our program director said at NSF was he's very happy to give us additional funding...because of the tremendous press the project has generated."

The Hub encourages active involvement from young people and has provided ways to engage them that align with the needs of both professors and students. One partner organization who was looking for a way to involve summer student interns in year-round activities noted that the Hub proved a valuable resource in their effort. The Hub helped another faculty member recruit students for a project.

Interviewees also noted that the Hub provides valuable resources for talking about science. One respondent described how the Hub provided science communication tools that have helped them better describe what they do using more accessible language. Also, the COVID Information Commons project and website helped a Hub stakeholder develop new communication tools for talking about the pandemic. They reported, *"One of my goals is outreach about what's happening on a larger scale, how we communicate about what's happening."*

The Hub has limited staff who are responsible for supporting many different stakeholders across the data science community, as well as a range of partner projects. Given the need for support on such a large scale, staff capacity can quickly become overwhelmed or overextended. Several interviewees noted the potential for tension exists between the desire to provide widespread support versus the need to keep efforts feasible. But they thought the Hub effectively navigates this balance currently and is realistic about the limits of its capacity. One interviewee acknowledged the process of identifying *"appropriate scope and resources"* for each of the Hub's initiatives as perhaps the greatest challenge to the Hub's success.

"It's gonna be important to keep on mission for a lot of these discrete initiatives of the Hub so there's a clear understanding of what the goals are for each initiative and project, how to get involved, and what is and is not 'in lane' for that particular program. I think that's hard for people like [current Hub staff] who are so inclusive...there's a tendency to want to be all things to all people, or to take on too much. But I think that they're both very clear about limitations in that regard."

Interviewees suggested that one way to accomplish this balance and ensure representativeness would be to develop additional strategies for community input, such as various ways to provide feedback via the website, or as part of community discussions following events such as lightning talks.

Recent Progress: Scale

Scale measures the extent and depth of change attributable to the Hub's activities, projects, and initiatives. In the first round of the evaluation, respondents talked about how the Hub's efforts are helping to advance the data science field and promote broader awareness about data science. Key findings from the second round of interviews build on those earlier findings:

- The Hub’s recent efforts have helped move the entity into a position where it can scale its reach and impact; and
- The Hub is continuing to grow by amplifying the work of its existing members and reaching new potential stakeholders.

The Hub is seen by its members as a place to accelerate collaboration – the experts involved in the network are considered “vetted,” which eliminates one of the potentially time-consuming steps of developing project or proposal teams. This past year has seen an unprecedented level of collaboration on the COVID Information Commons – an initiative that transcends the Northeast region and is indicative of current and future scale. In the words of one interviewee,

“I think these types of initiatives are just really needed right now because it’s such a pressing issue – the pandemic – and there are so many different people around the world working on this research. So, any way to bring them together and promote collaboration...is just really helpful. Without that, the research progress could be delayed.”

Building a long-lasting community requires reaching out to new potential stakeholders and engaging them in the Hub’s projects and network. Through the interview process, it became clear that those working with the Hub see a connection between its efforts to support diversity and inclusivity and the potential for scaled impact. They highlighted the diverse pool of seed fund awardees, which included those working outside of computer science disciplines and researchers at different stages of their careers. Interviewees also noted the Hub leadership’s deliberate focus on faculty dedicated to multi-disciplinarity as an example of its efforts to broaden the pool of participants in the community.

Interviewees highlighted students’ involvement in the Hub as crucial to its long-term growth. Several students have received stipends or scholarships from seed funding which has helped them stay involved and engaged with the Hub and other members. This two-tiered structure – where a researcher gets a grant and then re-distributes it to support student involvement – is evidence of the exponential potential of the Hub’s efforts. These funds are a boon especially for students who have to work while attending school and cannot take on unpaid research internships and projects.

Interviewees identified potential challenges to scale but did not find them insurmountable. As the Hub scales its efforts and impact, a question that remains is how best to capture seed fund-related successes without placing an undue burden relating to reporting on grantees. It will also be important to share successful efforts in a way that preserves the easy-to-navigate feel of the Hub’s website. Success stories are valuable for describing the Hub’s impact and benefit to the community, and interviewees noted the importance of thinking through the logistics of this process.

Another concern is the administrative challenge of processing and disbursing funds for the seed grants. The disbursement process is currently run through Columbia University but has been slow and inefficient – a marked departure from the responsiveness that has characterized communication coming directly from Hub staff.

Overall, interviews suggested that the potential for scale exists and the Hub’s efforts are increasingly likely to result in positive change, partly because so much work has been done to lay a solid foundation. Looking across the Hub and the projects and initiatives that members are

involved in, there is a breadth of activity involving a diverse community of stakeholders. Continuing to promote and amplify these efforts will be key to growing the Hub's impact and reach long term.

Discussion & Recommendations

Throughout the evaluation process, Knology received consistent feedback from the community about the Northeast Hub's support and engagement with the data science community. The Hub has played a critical role in bridging gaps between various sectors, creating opportunities for networking and broader collaboration, and providing valuable resources to members. Interviewees repeatedly called out the efforts of Hub's leadership and particularly the current Executive Director, Florence Hudson, as key to the progress that the Northeast Hub has made in the last year. Under the new leadership, prior efforts to connect and engage with the community have now been amplified and accelerated. Community partners have a clearer idea of how to engage with the Hub and what opportunities are available for them to do so. This is due to streamlined communication about Hub activities, as well as the updated strategic plan, both of which have strengthened the community's connection to the Hub and increased participation.

A recurrent theme was the recognition of the Hub's unique position in the data science community. Interviewees across both rounds of evaluation noted that the Hub's role cannot be adequately filled by government entities, academic institutions, or companies alone. As evaluators, Knology saw repeated evidence of the Hub's reach and influence, which far exceeded the bounds of the Northeast region, any one sector, or single academic disciplinary area. At the time of this report, the Northeast Hub's audience consists of individuals from all 50 states plus Puerto Rico and Washington D.C., as well as 14 countries. Through its efforts, the Hub has connected over 2,600 individuals spanning academia, non-profit, industry, and government.

This web of connections made through the Hub has resulted in funded projects and fostered collaboration across disciplines, institutions, and sectors. These partnerships translate into real-world projects with the potential to address important societal challenges. Several interviewees cited the COVID Information Commons as an example of applying data science research to practice for real-world impact. The CIC has reached over 1,100 individuals across 399 organizations since its July 2020 launch. Furthermore, Hub leadership have demonstrated that they are committed to listening to the community and implementing programs and activities that meet stakeholders' needs. For example, the COVID Information Commons Student Paper Challenge grew out of suggestions from students and has actively involved them in its design and launch.

Critical to the success of the Hub projects and collaborations is the Executive Team. Under the new leadership, the reputation of the Hub as a place for meaningful connection has continued to grow by leaps and bounds. Several Hub members specifically called out the efforts and commitment of the current Executive Director as crucial to facilitating interactions and building a strong network. Furthermore, more participants in the second round of interviews expressed greater satisfaction with their interactions with the Hub. Key to that satisfaction was much clearer communication from the Hub leadership regarding community activities and events, and the updates to its website that more clearly delineate its focus areas, its activities and projects, and its stated mission and vision for the community. Interviewees described a greater awareness about

opportunities for the community to participate, which has helped foster a stronger sense of belonging and connection.

Following the first round of the evaluation, the Hub implemented additional strategies for engaging with the data science community. This included increased participation in conferences and events that are frequented by members of the data science field. Community events like lightning talks provide opportunities for researchers to share their work and receive feedback from peers. These community-based events have been instrumental in fostering new partnerships, some of which have resulted in proposals for new projects. The first round of interviews showed a budding sense of community among members, a sentiment that had grown significantly by the second round of interviews as the Hub leadership became more intentionally committed to diversity and inclusivity.

Through initiatives such as seed grants and the Northeast Student Data Corps, the Hub has increased access to resources that support research and help train the next generation of data science researchers and professionals. Following round one of the evaluation, the Hub leadership made more concerted efforts to plug students into the Northeast Hub's ecosystem. As of December 2020, the Hub has reached 360 institutions of higher education, of which 25% are minority-serving institutions. The leadership team has also worked to lower the barriers to entry for the students who need it most. In 2020, the Hub's Seed Fund program awarded 19 seed grants to 16 institutions. Three of those awards were granted to minority-serving institutions.

Efforts have also included developing programs such as the Northeast Student Data Corps, which teaches data science fundamentals to students across the Northeast with a special focus on minority-serving institutions. During round two of the evaluation, there is stronger evidence of student involvement in Hub projects and in planning events and activities with the full backing of the Hub leadership. In addition to the CIC student paper challenge, there are opportunities for students to serve as co-PIs on seed fund grants. Students also get exposure from being involved with Hub leadership at prestigious conferences.

Overall, the evaluation demonstrated that the Hub is successfully meeting each of the four goals laid out in its strategic plan. It is clear that this is, in large part, due to the hard work and dedication of the Hub leadership. There are still some challenges to address as the Hub continues to grow – particularly around staff capacity – but it has a strong foundation on which to improve and grow with continued NSF support. In sum, analysis across both rounds of evaluation demonstrate that, building on successful efforts to date, the Hub has a critical future role to play in increasing data science capacity, access, and opportunity throughout the Northeast region and beyond.

Recommendations

The Hub has made a lot of changes in recent months that have substantially improved communication with members, drawing in a wider pool of stakeholders. The recommendations below highlight some of the current areas of effective practice that should be continued, as well as areas for improvement moving forward.

- Efforts to increase diversity and inclusivity have not gone unnoticed by the Hub's members. These efforts should be continued to encourage even greater participation from communities that are underrepresented in data science.

- Interviewees saw a lot of potential with the seed grant awards program – particularly for students experiencing financial hardship – but acknowledged that there is room for improvement regarding logistics. To scale the impact of this program, we recommend working towards greater efficiency in the disbursement process and ensuring better alignment between the timing of the awards and the academic calendar.
- Communication about the Hub’s activities and opportunities for member involvement has become much clearer. Current efforts should be sustained and built upon as the Hub grows its network. Interviewees suggested developing additional strategies for community input, knowing that the Hub is extremely responsive and welcomes feedback from community members.
- Capturing success stories is crucial for demonstrating the Hub’s value and benefit to the community. We recommend offering participants a range of ways to share their successes. Doing so will be mutually beneficial. For instance, when a data science researcher gets positive exposure that can be leveraged to advance their work, the Hub will be better able to speak to the outcomes resulting from the support they provide.

References

Ardalan, N., Norlander, R. J., Thomas, U. G., Nock, K., Brucker J. L., & Voiklis, J. (2020). Northeast Big Data Innovation Hub: Mid-Point Evaluation. Knology Publication #NSF.159.560.01. Knology.



Appendices

Appendix A: Round One Interview Protocols

Knology has been hired to help the NE Big Data Innovation Hub assess its impact. To do so, Knology will interview two groups of stakeholders: **project teams** and **community partners**.

The evaluation is designed to explore:

- **Reach / Engagement** measures the number and variety of communities affected by HUB activities and the number and variety of partnerships the HUB works with.
- **Integration** or **Interconnectedness** measures the extent to which HUB partnerships amplify and cross-pollinate HUB related activities to create economies of scale.
- **Impact** measures the perceived added benefit for stakeholders related to collaboration or integration with the Hubs, and the degree to which those who volunteer their time on Hub activities are satisfied with those efforts.
- **Representativeness** is the extent to which Hub activities align with and support the goals of partner and community projects.
- **Scale** is the extent and depth of change attributable to the activities, projects, and initiatives.

Text enclosed with brackets (i.e., [text]) represent explanatory notes for the interviewer, and the participant will not see or hear these notes. Examples include: [Prompt] or [Follow-up]

Round 1 Project Team Interview Instrument

Introductory script

Thanks so much for agreeing to speak with us today. My name is [Interviewer's name] and my colleague here is [insert transcriber's name]. As you saw from our initial email, we're with Knology, a social science research organization that has partnered with the Northeast Big Data Hub to help understand their current community engagement efforts - and how the Hub might evolve or continue current efforts to increase the positive impact for the Hub and community in the future. As we said in the email, this is the first of two half-hour conversations, with a follow-up planned for about one year from now.

We'd like you to be as honest as possible during this conversation. While the Hub knows who we are speaking with, we will not identify interviewees in our reporting, so they won't be able to match up people with their specific comments. [name of notetaker] will be taking notes, while the two of us talk. If it's ok with you, we'd also like to record the conversation. We won't share this recording with anyone, it's just to make sure we capture the conversation accurately. May I record?

[If yes, begin recording, if not transcribe only]

Do you have any questions before we begin?

Thanks!

Questions

For our record, can you say your name again and a sentence or two about who you are and the work you do.

[REACH/ENGAGEMENT] Can you tell us a little about your relationship with the Northeast Big Data Hub. Anything you think is important for me to know.

[Prompt] In what way do you interact? What value do/did you get out of these interactions?

[Prompt] Are you currently working on a project together or have plans to work together?

[Prompt] How did that work originate (e.g., out of a Spoke project)?

[INTEGRATION/INTERCONNECTEDNESS & REACH/ENGAGEMENT]

Have you made any useful professional connections through the Hub?

How did that connection result in other valuable interactions?

[Prompt] What value did you get out of this interaction? e.g., shared resources, maintain an ongoing dialogue, built a new collaboration or project, furthered your research?

[IMPACT] How do you feel the Hub supports or might support you or your work?

[Prompt: Are there any goals that they are helping you to achieve?]

[Prompt – As applicable] Does the Hub support translation or transfer of academic research into a real-life application? [Follow-up] What evidence do you have that this is occurring? [Interviewer asks about specific projects, depending on interviewee and the goal(s) of their project(s).]

[Prompt] What is your level of satisfaction with your relationship working with the Hub? (FH question - are we planning to give them a high, medium, low range, or ask them to just tell us in their words?) Can you tell me more about the reasons for your answer?

[IMPACT] What are the biggest challenges or obstacles, if any, getting in the way of your progress on Hub activities?

[SCALE] What are the biggest opportunities that you see for your work with the Hub in the future?

[Prompt] Can the Hub help mitigate the challenges you just described or support new opportunities? e.g., Specific skill-building? Resources? Collaboration?

[REPRESENTATIVENESS] Thinking about the diversity of work currently happening in your area of expertise, where do you feel the Hub should place its priorities to ensure they are supporting the diversity of programs in the field?

[IMPACT] What might be lost if an entity like the Hub didn't exist?

Conclusion

We are nearing the end of our time together. I'd just like to ask you one final question. Is there anything else that you think is important for me to know that we didn't already cover in today's conversation?

If you think of anything you'd like to add you're more than welcome to email us. Thanks again for your insights.

Round 1 Community Partner Interview Instrument

Introductory script

Thanks so much for agreeing to speak with us today. My name is [Interviewer's name] and my colleague here is [insert transcriber's name]. As you saw from our invitation email, we're with Knology, a social science research organization that has partnered with the Northeast Big Data Hub to help understand their current community engagement efforts - and who they might work with in the future. As we said in the email, this is the first of two half-hour conversations, with a follow-up planned for about one year from now.

We'd like you to be as honest as possible during this conversation. While the Hub knows who we are speaking with, we will not identify interviewees in our reporting, so they won't be able to match up people with their specific comments. [name of notetaker] will be taking notes, while the two of us talk. If it's ok with you, we'd also like to record the conversation. We won't share this recording with anyone, it's just to make sure we capture the conversation accurately. May I record?

[If yes, begin recording, if not transcribe only]

Do you have any questions before we begin?

Thanks!

Questions

For our record, can you say your name again and a sentence or two about who you are and the communities you work with relation to Big Data.

[REACH/ENGAGEMENT] Can you tell us a little about your relationship with the Northeast Big Data Hub. We realize this might be a new relationship, but we would like to hear how you learned about the Hub and what you've done together so far – if anything.

[REACH] Have you made any useful connections through the Hub?

[Prompt] Have you leveraged those connections in some way? e.g., engaged in a collaboration or program?

[IMPACT] How do you feel the Hub supports or might support you or your work? [Prompt: Are there any goals that they are helping you to achieve?]

[Prompt – As applicable] Does the Hub support translation or transfer of academic research into a real-life application? [Follow-up] What evidence do you have that this is occurring? [Interviewer asks about specific projects, depending on interviewee and the goal(s) of their project(s)]

[IMPACT] What is your level of satisfaction with your relationship working with the Hub?
What do you feel drives that level?

[IMPACT] What are the biggest challenges or obstacles you see for leveraging data in your field?

[SCALE] What are the biggest opportunities for leveraging data in your field that you see in the future?

[Prompt] Can the Hub help you achieve your goals or support new opportunities? e.g.,
Specific skill building? Resources?

[REPRESENTATIVENESS] Thinking about the diversity of work currently happening in your area of expertise, where do you feel the Hub should place its priorities to ensure they are supporting the broad diversity of programs in your field or the broader data science ecosystem?

[IMPACT & SCALE] What else could the Hub do to better support your long-term goals or vision in support of data initiatives?

[IMPACT] What might be lost if an entity like the Hub didn't exist?

Conclusion

We are nearing the end of our time together. I'd just like to ask you one final question. Is there anything else that you think is important for me to know that we didn't already cover in today's conversation?

If you think of anything you'd like to add you're more than welcome to email us. Thanks again for your insights.

Appendix B: Round Two Interview Protocol

Knology has been hired to help the Northeast Big Data Innovation Hub assess its impact. To do so, Knology interviewed two groups of stakeholders in May-June 2020: **project teams** and **community partners**. This first round of interviews resulted in a report summarizing findings and outlining recommendations for continued impact. The Hub leadership responded to the report and, together with evaluators, has decided to shift the evaluation approach, in part to reflect the vision, focus and activities of a new Executive Director, Florence Hudson, who assumed the role in May 2020. Instead of interviewing the same set of stakeholders again, Knology will interview data science researchers and practitioners who are **actively involved** with the Northeast Hub's current projects, some of which continue from earlier days and some of which are new initiatives not included in the original grant proposal. This approach will allow evaluators to gain deeper insight into timely collaborations – including a robust response to COVID-19 – rather than re-assess historical relationships that may no longer be as relevant. We note that given the Northeast Hub's new strategic plan and topical as well as regional focus areas, interviewees may live outside the geographic area and still collaborate with the Northeast Hub. For Round 2 we plan to conduct one-hour semi-structured interviews with 10-12 participants, to be identified by the Northeast Hub. In the case that Round 2 interviewees had also been interviewed in Round 1, researchers will return to the original data and ask questions that follow up on what interviewees said, while also asking about newer or more recent involvement with Hub initiatives.

Round 2 of the evaluation continues to explore:

- **Reach/Engagement** measures the number and variety of communities affected by Hub activities and the number and variety of partnerships the Hub works with.
- **Integration or Interconnectedness** measures the extent to which Hub partnerships amplify and cross-pollinate Hub related activities to create economies of scale.
- **Impact** measures the perceived added benefit for stakeholders related to collaboration or integration with the Hubs, and the degree to which those who volunteer their time on Hub activities are satisfied with those efforts.
- **Representativeness** is the extent to which Hub activities align with and support the goals of partner and community projects.
- **Scale** is the extent and depth of change attributable to the activities, projects, and initiatives.

Text enclosed with brackets (i.e., [text]) represent explanatory notes for the interviewer, and the participant will not see or hear these notes. Examples include: [Prompt] or [Follow-up]

Round 2 Interview Instrument

Introductory script

Thanks so much for agreeing to speak with us today. My name is [Interviewer's name]. As you might have gathered from the email exchange to set up this interview, I'm with Knology. Knology is a social science research organization that has partnered with the Northeast Big Data Innovation Hub to help understand their current community engagement efforts - and how the Hub might evolve or continue current efforts to increase the positive impact from the Hub for the community in the future.

[Alternate version of introductory script for repeat interviewees should acknowledge when that interview occurred and suggest that this one will proceed in a similar manner.]

We'd like you to be as honest as possible during this conversation. While the Hub knows who we are speaking with, we will not identify interviewees in our reporting, so they won't be able to match up people with their specific comments. If it's ok with you, I'd like to record the conversation so that I don't have to try to take notes while we speak. We won't share this recording with anyone outside of Knology, it's just to make sure we capture the conversation accurately. May I record?

[If yes, begin recording, if not take notes only]

Do you have any questions before we begin?

I do have some questions I'd like to ask you, as you are really the expert here. While I'm familiar with the Hub's many projects and initiatives, I'm most interested in understanding your experience with the Hub and the community. What we envision for these interviews is a flexible conversation that allows YOU to tell me what you think is important for me to know about how you work with the Hub. How does that sound?

Great, here we go.

Questions

For our record, can you say your name again and a sentence or two about who you are and the work you do.

[REACH/ENGAGEMENT] Can you tell us a little about your relationship with the Northeast Big Data Innovation Hub. Anything you think is important for me to know.

[Prompt] In what way do you interact? What value do/did you get out of these interactions?

[Prompt] Are you currently working on a project together or have plans to work together?

[Prompt] How did that work originate?

[Prompt] Was it clear to you how to get involved with the Hub?

[Prompt, as applicable] Are you familiar with [the Northeast Hub website / the joint Hub website / the COVID Information Commons website]? When you visit that site, do you know where to go to get the information you need? Has the website been helpful for you?

[Prompt] How do you learn or stay informed about Hub [and/or CIC] activities? If you wanted more information about one of the Hub's initiatives, do you feel you would know where to find that information? In other words, is information about what the Hub does accessible and clear?

[INTEGRATION/INTERCONNECTEDNESS & REACH/ENGAGEMENT]

Have you made any useful professional connections through the Northeast Hub?

How did that connection result in other valuable interactions?

[Prompt] What value did or do you get out of this interaction? e.g., shared resources, maintain an ongoing dialogue, built a new collaboration or project, furthered your research?

[Prompt - if the interviewee has multiple connections with the Hub] What was your initial contact with the Hub? How/why did you engage with additional activities since then?

[IMPACT] How do you feel the Hub supports or might support you or your work?

[Prompt] Are there any goals that they are helping you to achieve?

[Prompt] Do you feel like you have a clear understanding of what the Hub is / does?

[Prompt, as applicable] Does the Hub support translation or transfer of academic research into a real-life application? [Follow-up] What evidence do you have that this is occurring? [Interviewer asks about specific projects, depending on interviewee and the goal(s) of their project(s).]

[Prompt] What is your level of satisfaction with your relationship working with the Hub? You can use a scale from 1-10 or just describe your satisfaction. Can you tell me more about the reasons for your answer?

[Prompt, if familiar] Do you feel that the Northeast Hub's website would be a helpful way to promote the work you do? How would you envision this occurring?

[SCALE] What are the biggest opportunities that you see for your work with the Hub in the future?

How have you seen your work with the Hub grow or change over time?

[Insert initiative- or project-specific questions here based on what is known about the interviewee or the evaluation goals that person can best speak to.]

[IMPACT] What are the biggest challenges or obstacles, if any, that you anticipate might hinder your progress on activities with the Hub?

[Prompt] Can the Hub help mitigate the challenges you just described or support new opportunities? e.g., Specific skill-building? Resources? Collaboration?

[REPRESENTATIVENESS] Thinking about the diversity of work currently happening in your area of expertise, where do you feel the Hub should place its priorities to ensure they are supporting the diversity of programs in the field?

How well do the Hub activities you participate in align with or support your goals? You can use a scale from 1-10 or just describe this. Can you tell me more about the reasons for your answer?

[IMPACT] What might be lost if an entity like the Northeast Hub didn't exist?

Conclusion

We are nearing the end of our time together. I'd just like to ask you one final question. Is there anything else that you think is important for me to know that we didn't already cover in today's conversation?

If you think of anything you'd like to add you're more than welcome to email us. Thanks again for your insights.



Knology

Behaviors

Biosphere

Culture

Media

Wellness

Systems

Knology.org
info@knology.org

tel: (442) 222-8814
3630 Ocean Ranch Blvd.
Oceanside, CA 92056

tel: (347) 766-3399
40 Exchange Pl. Suite 1403
New York, NY 10005