National College Access Network

Building Connections. Advancing Equity. Promoting Success.

June 2017

Idea Incubator Series

Managing the Database Lifecycle and Ensuring Quality Data

Introduction

NCAN has heard repeatedly and consistently over the years that one of the best aspects of network membership is the ability to collaborate with and learn from other members. At our annual national conference, Spring Training events, and other convenings, the cross-pollination of ideas and resources helps make the entire membership stronger. In an effort to further this kind of transfer of ideas, in the summer of 2016 NCAN hosted a series of four Idea Incubators across the country, which brought members together to tackle tough questions in data and evaluation.

The objectives of these incubators were to develop a shared understanding of known research and best practices around the topic area; to generate new insights, questions and ideas to test in practice; and to identify actions to be taken and/or follow-up communication, as needed. Five to seven members came to each Idea Incubator. These two-day events were largely unstructured but were guided by a set of questions created by NCAN and bolstered by programs' individual interests. As discussions proceeded according to members' particular needs and levels of experience, broad themes emerged, as did shared needs for knowledge and additional resources.

The second of these Idea Incubators, "Managing the Database Lifecycle and Ensuring Quality Data," was hosted by College Forward in Austin, Texas. Readers should consider this document a hybrid between a white paper, a resource guide, and meeting minutes on this important topic, which has come into increasing focus in both the research literature and program practice in recent years.

Attendees identified a number of challenges: storing data across multiple systems rather than a centralized collection; collecting and sharing real-time data across different platforms; ensuring clean, standardized data; and obtaining "buy-in" from each member of the organization. Attendees also discussed structuring data points examining student outcomes and connecting campus services to these data, collecting programmatic data in addition to student-level data, facilitating internal site knowledge and distribution, and best practices for transitions to new data platforms.

Guiding Questions

The following questions guided the "Managing the Database Lifecycle and Ensuring Quality Data" Idea Incubator:
- By what measures do programs assess how well a database is meeting their needs?
- What are the different steps involved in the database procurement process? How can organizations effectively manage the database lifecycle from procurement to retirement?
- What are the green and red flags to consider when comparing database vendors?
- How can a program find a database that meets the needs of staff at various levels? What are some of those needs?
- What is data integrity and how does an organization know if it has it?

- How can an organization create, or improve, its policies and processes around data management?
- What are some common forms of user error related to data entry, data management, or data analysis? How can they be mitigated?
- How do organizations avoid duplicate data altogether and how do they handle it when they have it?
- What are some database and programmatic controls that can be built into a system to promote higher-quality data? What are some examples from the participating programs?

Organizational Data Strategy, Broadly

Programs wanting to ensure quality data must be thoughtful about which data they collect in the first place. "We want to collect any and everything about every single student," one incubator attendee explained. This is an example of one of the challenges faced by many NCAN member programs: if some data are good, then more data *must* be better. Another attendee pointed out that this is not necessarily true, and data collection, analysis, and usage should be tied to a strategic plan and, more broadly, an organization's mission and goals. Questions like "Who are we?" and "Who do we want to be as an organization?" are critical for programs to map out how they will effectively use data, while more narrow questions like "How often will we collect this?" and "Who will use this?" help to refine an organization's strategy.

A related best practice is to ensure that a program has a well-constructed logic model as well as an overlay of that logic model that considers data points and sources for each component of the logic model. NCAN's Common Measures may be useful here, and one attendee noted that their organization's use of these has helped them to get "a lot more focused on what we're collecting and why."

This idea of narrowing the data collected through a strategic planning process and identifying the metrics that point to programmatic outcomes had traction among attendees. "We're limited by the data we can collect, but we also want to have a strategy about why we collect this data. What are we going to use it for? We're starting to improve the collection piece, but we also need to improve the 'how are we using it' piece," one attendee explained. The merits of a data inventory are clear here. Organizations should ask themselves what they have on-hand, whether and how they use it, and what they learn from it. It is also critical to take a hard look at what prospective uses the data may have. There is a balance between stopping the collection of a data point that a program does not currently use but sees real potential for in the future, and stopping the collection of a data point that "maybe, might, someday, somehow, could be useful, possibly." The former is a good candidate for retention; the latter, less so.

The question of how to prioritize fields will vary by organization, but there was general agreement among the group that a variable that could be a predictive factor, or is an outcome, is more valuable than a process variable (e.g., student turned in a permission slip or waiver, especially if it's a historical variable from a past class). The idea here is

turning a critical eye toward the question of "Do we need to keep this variable?" You say you need it now, but will you need it tomorrow?

Attendees noted that funders have been drivers of data collection and have brought to the forefront tracking factors of academic success (more on funders shortly). Although attendees' organizations track data in different ways, there was discussion around the need to *use* the data once it is collected. Attendees discussed using data around both broad programmatic outcomes and day-to-day activities that lead to those outcomes. Attendees noted that databases can be research tools and set an organization up for a future program evaluation, but questions persisted around the kinds of data practices that set an organization along this path.

Despite the clear need for organizations to be strategic about the data they collect and analyze, attendees raised the issue of funders as drivers of data collection. Funders often require programs to collect and report on various metrics, but these metrics may not always fit into an organization's data plan. Whether it is a new metric for an organization or a metric reported in a new way that is at odds with a program's typical reporting, attendees noted that funders sometimes ask for data that is only actually useful for the funders.

One attendee explained:

> We're trying to negotiate what's feasible and usable for us or for our funder. What's in the grant doesn't align with our practice and what they need. We ask them for best practices for how their other grantees engage in those practices. How can we adhere to funders' needs while still trying to passively suggest other ways in which to modify some of their practices and asks of us?

Another attendee noted, "In the last couple of years, we put a stake in the ground to not undergo any new initiatives. A lot of the strict things that we had to do…for grants have disappeared. It was something that as part of the grant we needed to do, and we're moving away from that model." Pushing back on funders in this area is understandably difficult and requires an honest working relationship between both parties. Funders necessarily want to see the outcomes of the projects and organizations that they are funding, and organizations want to be able to devote as much time as possible to their projects and broader organizational mission. Much of the data a program collects is for grant reporting and funders who understandably desire to see the outcomes of the projects and organizations they are funding.

Promoting Data Buy-In

This Idea Incubator segment revolved around staff "buy-in" and striking a balance for staff between data and non-data activities. Attendees reported that their organizations' data analysts or, more generically, "data people" are the few who manage the bulk of the data within their organizations. Despite the concentration of data management in the hands of a few, promoting full staff buy-in around knowing how to track and analyze

data increases staff data knowledge and use drastically. "With the gross tonnage of data collected, how do you find the balance between data entry and programmatic activities?" one attendee asked. "What are the must-have data points for predicting student success?"

To this end, attendees' organizations promote staff buy-in around data in a number of ways. First, organizations can signal through the hiring, onboarding, and ongoing training processes that data are a priority. One attendee discussed how they "convey the importance of data on the front side. For us, it's naming it" by including a direct percentage of time spent using data in job descriptions (e.g., X percent of job responsibilities will include data administration, entry, training, coaching, etc.). Once staff are hired, organizations often do more of the data use training in the first couple of weeks to emphasize why they use data in the first place and cement it during orientation as an organizational priority. Following up orientation with staggered data training across the year as students progress and new analyses are needed is also a useful activity. To promote familiarity with data and increase a feeling of accountability for it, some organizations distribute data duties through activities like filling in missing data, data cleaning, and taking data analysis back to their respective teams.

Another way to promote broader buy-in is to be more thoughtful about the signals sent around who is responsible for data. Organizations with designated "data people" run the risk of "non-data" people feeling excluded. "We should integrate more practitioners into the data collecting role by making the data more useful to those who are not the data manger," one attendee noted. Wherever possible, reports and visualizations should be relevant to the frontline staff who collected the data. Being able to see the results of data and its findings make data more engaging to all staff members, not just the data managers. Knowing how data benefits them and how the use of data makes their jobs easier is also a strong selling point. As one attendee noted, "Fun data makes for good data!"

For more on data buy-in, consult [Driving Toward Program Improvement: Principles & Practices for Getting Started with Data](#).

Data Training

Some attendees identified earning buy-in and providing training for staff who will only be with their programs for a year (e.g., AmeriCorps members) as their biggest challenge. These temporary or contracted workers are performing a wide range of duties around data, but are often unable to make a connection to the work to which they are contributing because of a lack of understanding around data use and practice. The upshot is that the temporary staff model generates a regular need for staff training, which provides an opportunity to improve the quality of that training.

Another challenge associated with temporary staff includes the need for monthly training because of varying onboarding windows. As a remedy, some members have developed recorded training modules, which seemed to work more favorably despite leaving the training requirement in the hands of the trainee. Screencast-O-Matic is an example of a platform that can be used to record virtual training modules where trainees

can learn at a comfortable, self-directed pace. From there, students can take quizzes via other platforms like Google Forms to verify that they retained the lessons from the videos.

One attendee advised spacing trainings throughout the academic year. "Don't frontload them," they said. "We do more of the 'why data' lessons in the first couple of weeks. That 'why' piece was crucial to us. 'Why do you think that we do data?' we ask." Spacing out the trainings helps programs check in at various points with staff to determine if they need a refresher on something, or are unclear about anything that can be corrected at a future training. It does an organization little good to ask outgoing staff retroactively what they wish they had been told earlier.

Beyond initial and scheduled training, attendees also had ideas about how to make frontline staff's feelings toward and experiences with using the database, positive. Ideally, staff will come to see the organization's database for what it is: a shared asset that everyone contributes to and benefits from. Unfortunately, there can be some resistance in getting there. "I just feel like people are so bored during the training," observed one attendee. "A lot of people don't come from a data background; 'I'm here to talk to these kids, I'm not here to type about these kids,' they will say. But they are interested to see how data could work for them." To excite the staff, one attendee put temporary staff into data teams where they can clean and fill in missing data. Another noted, "We have parties around the harder things to try to make a community space where people can be happy or sad together. We designate a set amount of time to sit together and review data." Those parties can help to generate new content for end users that demonstrates value for them, e.g., "Oh, that was a heavy lift, but at least we are learning things." Ultimately, staff should feel that "their data is really cool or important, even if it is with interim data releases," one attendee concluded.

Attendees also noted that frontline staff offer valuable insights into which data are collected, used, or missing. These staff know what is missing, can take stock of what they are or are not using, and discuss pros and cons from their perspectives. This deepens their ownership and voice in the data process and is a much more positive contrast to staff feeling like, "We're using it because you're telling me we need to collect it."

Despite best efforts during training, sometimes staff still face challenges or do not buy in to data management and use. Attendees discussed how to handle these challenges. One suggestion was to sit down with staff and map out the processes they use around data, i.e., going step by step through their workflow around key data tasks, to find out why they are not using databases and to help them find a solution. Attendees also recommended investing time in documentation of the data platform in the form of codebooks and manuals. "We create codebooks of how we are expecting districts to collect certain [key performance indicators]," one attendee noted. "We have also gone through a series of data inventories….We want examples and auditing protocols." There are myriad benefits to this documentation: It leaves a record in case of staff turnover, it provides a reference accessible to anyone, and it helps to provide cover for correcting staff mistakes by referring them to the guide.

Ensuring Quality Data

Clearly delineating staff data reporting responsibilities for different levels ensures that everyone has *some* fields that are important to him or her and promotes data quality. Perhaps even more importantly, these different responsibilities help to shape the questions asked and actions taken by staff at different levels. For example, executives might examine the organization's direction by seeing how many students have been contacted in a given quarter. Managers might view frontline staff and ask questions like, "Of those who aren't getting meeting their goals, how can I help them improve?" Frontline staff may look at student-level data and ask, "Who are my red-flag students and what do I need to do to help them improve?" Data structures and reports can also guide daily activities. For example, if an organization's focus of the month is ACT registration, noting which students have had that discussion with their advisor can drive a report that shows which students advisors need to meet with in a given week. Note that the timeframes are often different for various staffing levels. Frontline staff need to have data inform today and tomorrow, but a leadership team might want data to inform the next three to five years.

One attendee noted that they successfully promoted quality assurance and control by helping staff to understand that the data they are collecting is integrated into individual and organizational performance. To that end, the organization maintains multiple levels of key performance indicators (KPIs):

- Level 1 Indicators, by site (managed by Executive Director): Linked to national programmatic goals and include high school retention and college acceptance, matriculation, retention, and completion.
- Level 2 and 3 Indicators (managed by Program Manager): Include college application rate and FAFSA completion rate.
- Level 4 Indicators (managed by individual staff): Include individual performance measurements. "If [staff] can see where the data points plug into around where they're going to be reviewed, it helps."

Attendees also engaged in a long discussion around their particular quality assurance (QA) and quality control (QC) procedures. QA and QC are related; the former includes verification and auditing of data, while the latter relates to the processes and actions that go into providing accurate data in the first place. One attendee noted that they engage in "heavy data auditing" semiannually. The first audit checks for students' demographic and other characteristics, and is followed up by monthly audits to ensure that advisors fill in gaps in the data. The second audit happens before September, when grade promotions are occurring. Each data point in a student's profile is verified before they move on to the next grade, and fields where data are missing are flagged.

Another attendee advised, "Do a complete audit, not just random selection," but the level of burden required for a comprehensive audit may be too much for some programs. Although another member suggested that random selection can be valuable when it comes in the form of spot checks on data entry and quality, e.g., "Have you really only interacted with five students in the past six months or are you behind on logging your interactions?" Programs that log interaction data between students and

advisors found that getting this type of data logged in a system was the biggest obstacle, especially compared to relatively more stable data like student demographics.

"Staff are not going to leave on good terms with us unless we have data in our must-have fields," said one attendee, who went on to explain that in the past the organization had staff members holding out until the last two days before a deadline to enter their data. "What people get in the habit of doing is maintaining their own spreadsheets … We held ourselves to the premise, 'If it's not in Salesforce, it doesn't exist.'" This organization relies so much on reports and dashboards, especially during meetings, that putting the emphasis on data integrity and keeping data updated were critical. Part of that process is making sure that users have the ability to edit the appropriate records. "We think a lot about credential settings for users. Which users really need the ability to create, read, edit, or delete records [in our system]? If you are thoughtful about that, you can prevent a lot of data integrity issues downstream." Organizations, during their data inventory process, should map out which kinds of data and records their system handles and which kinds of users should have permissions to create, modify, or delete those records.

For more on ensuring quality data, consult [Roadmap for Tracking Your Student Results: College Access/Success Data & Systems](.).

Data Integration and Real-Time Data

Data integration, i.e., combining multiple data systems into one data system, was a prominent topic. Programs often use different systems to meet the full range of their program's needs. Multiple data integration challenges arose. In the college access and success field there are a number of "specialized" platforms that serve specific roles, e.g., preventing summer melt, preparing students for matriculation, or texting. How can these programs interact with more traditional client relationship management (CRM) platforms? To what extent should programs be concerned about repetitive data collection between multiple overlapping platforms? Although attendees expressed interest in wanting to connect multiple platforms, they also expressed concern about the extent to which data integration would call for manual data entry, which would be burdensome for organizations with large data sets. Organizations need to consider their programmatic logic model as well as the data overlaps and interactions between potential systems. They also need to work with platform vendors to discuss the level of effort required to integrate multiple data systems.

The above data integration strategy may be particularly difficult for some organizations, especially intermediary ones that are working with and sharing data across sites. Attendees discussed the challenges between "segmented and real-time data." Data-sharing agreements with partners, data integration into multiple systems, and working with outdated data all present challenges to organizations. Organizations are often unable to share their data because there is no centralized platform across all partnerships, which makes it more challenging to combine data effectively. There are technical and logistical challenges here. One attendee noted that because their partners are using different platforms, the organizations share data by exporting and sharing .csv files (a commonly available format across spreadsheet programs like Excel) and uploading them to different systems. The standing question was whether there are

application program interfaces (APIs) that can connect platforms to each other to provide real-time data. That level of technical knowledge eluded attendees (and likely eludes most NCAN members) but warrants further investigation and potential contracting to resolve.

Conclusion

Many participants realized they were not as ready to migrate and/or integrate their data systems as they previously thought, either because of their data's cleanliness or their platform's capacity. Despite this, the Idea Incubator included many takeaways regarding improving data's usefulness. New insights ranged from data talking points with non-data managers, data life cycle and relevancy, promoting staff ownership and voice in the data process, scaling data, and making data as user-friendly and simple as possible.

Remaining questions or needed resources out of this Idea Incubator included:

- A way to share report or data templates between organizations; same for different types of agreements: state-level, school district, other parties
- More case studies of different data platforms with testimonials from members (coming soon!)
- Examples of interesting reports or analysis from different places, perhaps even via a learning community for sharing internal and external analyses
- A "from start to sustainability" guide – a resource that talks about planning things at each stage of data management
- A Data and Evaluation Learning Community: a place learn terminology and techniques, especially for those without that background (in progress!)
- A developer guide for connecting APIs – how to connect data systems

During the closing of the Idea Incubator, participants offered their advice for other programs interested in database management:

- "It is really easy to assume technology will fix things, but the advice I would give is to hold the mirror up first, to evaluate processes and the data you're collecting, why, and do that part first. We need to evaluate our processes before we migrate into a new system."
- "Identifying what your organization's data culture is and where it is, and see if there's something there that needs to be addressed before you turn your eye toward a revised or new database."
- "If you are thinking about switching to a new platform, talk to your partners about what they're using. Get on the same page in terms of what's possible. Engage in some long-term strategic planning, especially if you will be engaging in data sharing with them. If the systems are not or will not be communicating in real time, think about data-sharing agreements. Make sure your partners are meeting your expectations and that you are meeting theirs."
- "Focus on building a culture of data sharing and transparency. Even if it is 30 minutes a day, it significantly impacts the bigger picture."

NCAN would like to thank all of the attendees for their candor during this Idea Incubator, as well as College Forward for being a phenomenal host for this event. Stay tuned for future briefs documenting the rest of the Idea Incubator series.