

Thoughts on the Organizational Structure of the Nationwide Data Initiative

Michael Holland
Center for Urban Science & Progress, New York University

Introduction

The Nationwide Data Initiative's goal is to build local capacity for evidence-based policymaking and evaluate the effectiveness of programs implementing evidence-based policies. My comments are made with the assumption that the Initiative Organization (IO) is intended to be the vehicle whereby private foundations channel support for the establishment of data centers that house linked administrative data across government units.

The IO will be responsible for three tasks:

- **Grant making:** The primary task here is the running of a policy-focused competition (RFP development, selection/learning phase, final project selection) as outlined in the NDI document.
- **Awardee support and oversight:** The NDI document correctly tasks the organization with fostering collaboration and lesson sharing among the data centers that are established with initiative funding. Given that the initiative aspires to produce data centers with sustainable business models at its conclusion, these infrastructure projects also require a robust oversight effort tracking awardee's cost, scope and schedule progress.
- **Promulgation of Replicable Data Strategies and Data Standards:** The NDI document tasks the IO with the creation and support of a Standards Consortium, including managing relationships with state and local governments, creating networks across awardees to facilitate learning by the evidence-based policy and program management community, and providing technical assistance across projects.

My input on the following organizational elements follows: focus of the initiative, incentives for data providers, governance structure, funding structure, timeline, measure of success, and staffing structures. Regarding the position of the Executive Director and the infrastructure build, the most useful thing I can do is endorse the advice given by Robert Grossman and David Kendrick in their white papers.

Focus of the Initiative

The emphasis on organizing individual data centers around distinct policy communities is correct. The proposed initial focus on improving economic mobility will integrate data across social services delivery, healthcare delivery, and education. These initial areas capture some of the most significant funding flows from the federal government to state and local governments.¹ The relatively high degree of federal administrative conditions attached to these grants (typically involving the imposition of federal standards for

planning, project selection, fiscal management, administrative organization, and performance) should generate extensive and reasonably comparable data across multiple jurisdictions.

Incentives for data providers

Program staff within agencies and non-profit data donors are going to be most strongly motivated by actionable operational insights from the provision of their data. Academics and policymakers are going to be most strongly motivated by the opportunity to demonstrate their technologies, methodological innovations, or assess the value of new program and policy designs. The potential mismatch in the time horizon for realizing these ambitions presents a tension for the data centers.

The IO should consider funding scholarships for executive education programs that build public sector analytics capacity, as well as follow-on consultants who can work with users on preparatory work (data mapping, meta-data cataloging) or rapid, short-term analyses while the data centers stand themselves up. Quick, relatively simple wins help with momentum and help justify the continued participation of public sector partners who will always have large numbers of other demands for their time and resources. There is a non-trivial coordination overhead with this approach, so the IO will need to allocate funding to make it work. This “community seeding” strategy is not uncommon with some of the major scientific user facilities I’m familiar with.

Governance structure

The Governance Board should have balanced representation from the private foundations funding the NDI, the data centers’ user community, individuals who can span the scholarly and policy communities, and board members with technical expertise in building and operating data enclaves.

IO board seats will create the opportunity for the various foundation program managers to maintain a common understanding of NDI performance and continued commitment to the initiative’s goals.² Leaders from within the immediate user community provide an insistent pressure for results from the data centers. Since the scope of the policy and program user communities should expand over time, a few additional members representing the interests of the future user communities will keep the board connected to the broadest opportunities this data revolution poses for the public sector and cognizant of its most serious barriers. Ideal candidates for this ultimate user role are the visionary Chief Data Officers from cities who are successfully building their internal capacity to reuse data across organizational silos to drive policy development, program execution, and program evaluation. A couple of members from think tanks that span the scholarly and policy communities can, at their best, provide a perspective that spans and integrates the interests of the other board members.

No more than one technical expert on the board should be expected to serve in a representational role as the collective voice of the data centers supported by the initiative. If the total number of centers is small, it is preferable to have that *ex officio* role rotate rather than put all center directors on the board. Cases I’ve seen where all center directors

supported by a program sit on a board or advisory committee leads to situations where center directors feel compelled to prosecute issues at board meetings that are better left to separate discussions with the Executive Director. Lastly, if the Standards Consortium is to be a major programmatic effort of the IO, a member with expertise in the development of voluntary consensus standards would be a prudent addition. National Institute of Standards and Technology (NIST) staff frequently serve in this capacity on a wide variety of standards organizations, and NIST has a growing interest in data standards. Federal statistical agencies may have the potential to provide this expertise.

Most importantly, the Governance Board must have real authorities, otherwise it is merely an external advisory board. In the early stages of defining the responsibilities of the Executive Director, the authorities of the board also need to be clearly defined and coordinated with those of the Executive Director. For example, board sign-off at the start of the most significant stages of the competition may be appropriate, e.g., final language of the RFP and agreement on that stage's evaluation rubric.

My comments relevant to the technical/scientific board are included below with my comments on the IO staffing structure.

Funding Structure

I assume IO is a private pass-through foundation, so my comments relevant to funding of IO are included in the Staffing Structure section where I believe they are more germane. Here I address the structure of funding awarded by IO.

For the NDI to succeed, the IO needs to support the development of both the evidence-based policy/program community as well as the administrative data center community. Support for projects using linked administrative data should be separately funded from, but coordinated with, support for development of the data centers. Co-mingling funding for the user projects with funds for the data center development has the potential to encourage an inward focus on the part of the data centers – or even worse reduce the center awardee into little more than a micro-granting entity.³ Assuming that project awards will include funding for data center user fees, separate IO support for data center ops can ramp down as the centers mature and their user base broadens beyond the initial projects seeded by NDI funding.⁴

I know our brief is not to comment on the data centers, but I believe the IO must pay special attention to ensure that each data center includes one specific skill set, that of the data center research scientist. The skills needed to stand up and operate the data center are distinct from those of the user. By analogy to the x-ray synchrotron light source user facilities, the people who build and operate the accelerator complex and the beamline instrumentation are distinct (and somewhat more homogeneous) research communities compared to their incredibly diverse user communities, and I expect the same to be true in the case of these data facilities. Analogous to the beamline scientists at the light source, data center research scientists serve as the interface between facility and user community. Only a small number are needed in each data center, but they are essential if the data center is to couple effectively to its user base. The role “ISR-tenured” research scientists at

the Institute for Social Research may provide more detailed guidance for the design of these positions.

An essential area of technical assistance the IO needs to develop is a scholarly legal community that can tackle the well-known problems facing administrative data sharing that arise from uncertainty and the lack of explicit permissions in the rules defining data reuse. Information privacy law traditionally governs the collection of data, with a focus on notice and consent, or on the prevention of unauthorized data disclosures. Paul Ohm argues that few laws govern what can be done with data once legitimately collected.⁵ The few use restrictions that do exist build on the Fair Information Practice Principles (FIPPs) that are interpreted to restrict use to the purpose for which data was collected. The level of expertise required to develop the robust legal templates applicable in multiple jurisdictions is beyond the expertise any single data center team can or should muster. The ultimate goal for the legal scholarship should be a coordinated set of model legislation and data sharing templates that provide a solid foundation for the safe reuse of administrative data. These are tasks the IO should separately fund, coordinate, and integrate with the data center teams.

Timeline

The NDI document describes the competition stages as developing the request for proposals, initial selection and learning/development phase, final project selection, and collaboration and lesson sharing during the build phase. This is essentially correct.

During the institutional establishment stage, when the program goals are being sharpened and the competition is being designed, an essential task is workshopping the goals, competition design, and performance metrics with experts, potential applicants, and informed parties who are unlikely to compete but are likely to follow progress of the NDI and benefit from its outputs. That public vetting, which slows launch, is invaluable in making sure the IO is as well informed as possible about the genuine benefits and pitfalls of the program. Two rounds of workshops would be ideal (NDI version 1.0 > workshop > version 1.1 > workshop > version 1.2. > publish the RFP, version 2.0). The second workshop allows the IO to signal to potential applicants its evaluation criteria for each element of the proposals.

Once the RFP is on the street, the IO should offer very small grants (travel, meeting expenses) to allow interested teams to workshop their roadmaps for data center development. This should signal that the IO wants proposals that are not only serious in terms of their scientific content, but that the teams have taken the time to develop reasonably detailed software engineering project plans.

The initial selection and learning/development phase correctly identifies the need for a moderate-sized planning grant. That will allow teams to remedy any shortcomings identified in their technical roadmap, sharpen their scientific uses cases, and at least commit to paper a conceptual business model they would work toward if their proposal was selected.

Measure of success

In the near term, the NDI can take inspiration from the XPRIZE Foundation, an organization that offers inducement prizes for solutions to bold technical challenges. Their competitions must:

- have a bold and audacious goal;
- target market failure;
- define the problem vs. the solution;
- be audacious, but achievable;
- be winnable by a small team;
- be achievable in a reasonable time frame;
- be judged against clear, objective and simple rules;
- be telegenic and easy to convey;
- be leveragable;
- drive investment;
- create "back end" business; and
- provide vision and hope.⁶

Although the NDI awards are not inducement prizes, the NDI goals resonate strongly with virtually every one of the XPRIZE criteria. Adapting some of these, like the “target market failure,” shouldn’t be complicated, e.g., “target information exchange failures.”

The challenge for the Executive Director during the launch of the NDI will be to drive a conversation among the three main stakeholder communities (the public sector, researchers, and sponsors) to generate a common understanding of specific metrics that would follow from these criteria in the NDI context.

In the long term, the NDI should be judged on whether approaches it develops to integrate data across the initial focus of health, education and employment outcomes can find broader applicability across other core local government service areas. Coherent data pools should be possible in areas defined by commonalities of activities, legal and regulatory frameworks, and the flows of funding (and reporting requirements) from the federal government through the states to local governments. In time, the NDI should help foster a number of coherent public-sector data strategies; examples could include code-enforcement,⁷ licensing, public safety, or public infrastructure operations (e.g., roads, water systems, sanitation, parks).⁸ Given that real-time condition and use data on the built environment is increasingly pervasive (i.e., the internet of things), this is an opportunity to bring routinely unobserved conditions about the environment or built infrastructure into analyses of behavior (e.g., educational attainment of poor kids may be more influenced by the fact that schools with high populations of poor kids are more likely to be built in noisy, distracting locations than schools with high populations of more affluent kids).

IO Staffing structures

The staffing structure should parallel the IO’s tasks, which I interpret as grant making, project oversight, and the operations of a Standards Consortium. As described in the NDI

document, the IO sounds like a private pass-through foundation. Clearly, the IO staff must control all intellectual components involved with designing and executing the initiative, but a question worth discussing is whether the IO should build all the financial, HR and accounting machinery necessary to support the grant-making functions in house. The IO would likely be better served by relying upon one of the supporting private foundations for these back-office services – or establishment of an NDI Secretariat should be competitively awarded to an institution that would not compete for a subsequent data center as the first stage of the NDI.

Given the technical and legal complexity of the goals of the initiative, a more muscular version of traditional foundation program manager oversight is required. An IO staffer with expertise in project management is essential (specific expertise with software engineering/data enclave environments is highly advisable). That staffer should carry out a version of the regular external technical, cost, schedule and management peer reviews common in major research facilities construction projects adapted to the data center context. Good models have been developed by DOE's Office of Project Assessment⁹ or NSF's Large Facilities Office.¹⁰ These review panels are populated by a mix of practitioners and consultants, whose expertise will change dynamically over the life of the data center's development arc.

The IO should look to NIST's role in the Green Button Initiative¹¹ as a model for work of the Standards Consortium, which needs to develop protocols and model standards that are flexible, uniform, and technology neutral. The IO's ambition should be to include a broader group of stakeholders in the development of these standards than simply the subset of projects and data centers funded through the main NDI funding lines. As Box 2 of the January 23 *Preface and Roadmap* document illustrates, there is a larger community that the standards development process should seek to engage. Doing so will help ensure that the NDI-funded efforts integrate with and advance what exists currently.

It is not clear to me that staffing the IO with software engineers who develop data standards centrally for the individual data centers, as reflected in the NDI document's notional budget, makes sense. A Standards Consortium requires significant level of effort. It may be more appropriate to staff up the IO with a couple of engineers who have experience in setting consensus data standards and provision them with the resources necessary to support technical task forces and standards development workshops and provide travel stipends for select non-awardee participants. (Absent obvious commercial or near-commercial products, few organizations or public agencies would be in a position to justify self-funded travel for standards development.)

The IO can look to the Georgetown University Energy Prize as a good model for outreach and communication strategy.¹² They've consistently received local and national media attention for participating cities over the 3-year life of the competition.

Michael Holland is the Executive Director of Center for Urban Science & Progress at NYU. He previously worked as Senior Advisor and Staff Director in the Office of the Under Secretary for Science at the U.S. Department of Energy. Before that, Mike oversaw DOE's Office of Science

for a decade as a program examiner in the White House Office of Management & Budget, as a policy analyst in the White House Office of Science & Technology Policy, and as a staff member on the U.S. House of Representative's Committee on Science. He is a member of NSF's Business and Operations Advisory Committee and NIST's Smart Grid Advisory Committee.

-
- ¹ According to a Congressional Research Service analysis, FY2018 health care is anticipated to account for more than half of total outlays for federal grants to state and local governments (an estimated \$ 432.5 billion in FY2018, or 61.5% of the total), followed by income security (\$107.1 billion, or 15.2%; examples include TANF, LIHEAP, child nutrition), transportation (\$61.5 billion, or 8.7%), education, training, employment, and social services (\$59.5 billion, or 8.5%), community and regional development (\$16.5 billion, or 2.4 %), and all other (\$26.3 billion, or 3.7%). *Source:* Dilger, Robert Jay. "Federal Grants to State and Local Governments: A Historical Perspective on Contemporary Issues." Washington, DC: Congressional Research Service, R40638 (June 22, 2017).
 - ² Callen, J., Klein, A., & Tinkelman, D. (2003). Board composition, committees, and organizational efficiency: The case for nonprofits. *Nonprofit and Voluntary Sector Quarterly*, 32, 493-520. doi:10.1177/0899764003257462. This paper suggests there is a statistical association between organizational effectiveness and the presence of individuals considered to be major donors on the board of directors.
 - ³ Each of DOE's Energy Innovation Hub awards were funded at ~\$25M/year. Many fell victim to this micro-granting behavior rather than taking a disciplined approach to rapidly translating research into early-stage technology ready for commercialization. The most notable exception was the Nuclear Energy Modeling and Simulation Hub, which was able to build on the mode of developing community simulation code pioneered by the SciDAC program.
 - ⁴ While not so clearly applicable to facilities such as data centers with low relatively low fixed costs compared to the total cost of any single user group, the National Academies' *Cooperative Stewardship* report is a good analysis the potential pitfalls of user fees. See National Research Council. 1999. *Cooperative Stewardship: Managing the Nation's Multidisciplinary User Facilities for Research with Synchrotron Radiation, Neutrons, and High Magnetic Fields*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/9705>.
 - ⁵ Ohm, Paul. "Changing the rules: general principles for data use and analysis." *Privacy, Big Data, and the Public Good: Frameworks for Engagement* (2014): 96.
 - ⁶ <https://www.xprize.org/about/what-is-an-xprize>
 - ⁷ NYC DataBridge, which enables integration of data from agencies with enforcement authorities appealable to the Environmental Control Board is one example of a natural data pool. See *NYC by the Numbers. Annual Report – 2013*. Michael R. Bloomberg. Mayor. City of New York. Michael Flowers. Chief Analytics Officer. Mayor's Office of Data Analytics. December 2013. https://civicio.files.wordpress.com/2016/07/annual_report_2013.pdf.
 - ⁸ While the privacy protections related to personally-identifiable information (PII) such as HIPAA or FERPA are familiar, other statutory and regulatory frameworks, such as 6CFR 29, create Protected Critical Infrastructure Information (PCII) for a number of state and local physical infrastructure assets. A similar case can be made that that treasure troves of data from existing program are stove-piped, underused, and poorly maintained in these more engineering or operations-oriented areas as the NDI has made for human services.
 - ⁹ See *DOE/SC Independent Project Review Process* (January 2012) https://science.energy.gov/~media/opa/pdf/processes-and-procedures/1201_Review_Process.pdf
 - ¹⁰ See §4.5.3.2 of *Large Facilities Manual: NSF 17-066* (March 2017). <https://www.nsf.gov/pubs/2017/nsf17066/nsf17066.pdf>
 - ¹¹ Dr. David Wollman, NIST Smart Grid Program, *An Introduction to Green Button*. Available at <https://energy.gov/downloads/nist-green-button-presentation>
 - ¹² <https://gupep.org/category/media/>.