

## Meeting 17 Notes and Actions

May 20, 2022

**Next Meeting:** ACDEB Meeting 18: July 22, 2022

### Meeting Agenda:

1. Meeting Introduction, Agenda Review, and Reminders
  2. Recommendations Summary with Committee Discussion
  3. Subcommittee Report: Governance, Transparency, and Accountability with Committee Discussion
  4. Subcommittee Report: Other Services/Capacity-Building Opportunities with Committee Discussion
  5. Subcommittee Report: Technical Infrastructure with Committee Discussion
  6. Subcommittee Report: Government Data for Evidence Building with Committee Discussion
  7. Subcommittee Report: Legislation and Regulations with Committee Discussion
  8. Piecing It Together: Scenarios
  9. Next Steps and Action Items with Committee Discussion
- I. Meeting Introduction, Agenda Review, and Reminders** – Emilda Rivers, ACDEB Chair and Alyssa Holdren, ACDEB Designated Federal Officer (DFO)
- a. Welcome
    - i. Meeting 17 is an opportunity to discuss the ongoing work of subcommittees as they gather information, develop findings, and draft recommendations for Year 2 Report
    - ii. Will introduce recommendations that have received initial approval at the full Committee level: 1) recommendation for organizational structure for the national secure data service (NSDS) and 2) recommendations around communications  
*Note: Recommendations are drafts and subject to revision but represent a major milestone for the Committee's work*
    - iii. Will touch on the Committee's ideas around governance and technical infrastructure related to NSDS data capabilities to help lay building blocks for recommendations to be presented at July meeting

- iv. Will hear subcommittee reports on other activities beyond recommendations at the full Committee level
  - v. Finally, will touch on Committee's roadmap and next steps driving toward Year 2 Report with ambitious goal of completing recommendations by July meeting
- b. Introduction of New Chief Statistician – Dominic Mancini, OMB Office of Information and Regulatory Affairs
- i. In January meeting:
    - 1. Discussed progress toward addressing ACDEB's recommendations from Year 1 Report
    - 2. Began iterative engagement with this group, which continues today in this meeting
    - 3. Talked about evolving vision for statistical system, operating as seamless system of trusted data stewards to enable greater evidence building
  - ii. Want to make clear there is a framework in which system thinks about an NSDS but don't have many or any answers, so looking forward to recommendations of this Committee. ACDEB has been very welcoming with OMB/ICSP conversations, and OMB/ICSP is still learning, so really depending on recommendations of this Committee to inform thinking
  - iii. Since January:
    - 1. FY 23 budget was pending but has now gone out; contains funding for critical initiatives to advance evidence building that align with this Committee; includes funding for early work on NSDS; hasn't yet been adopted but is reflection of Administration prioritizing work of statistical system and data service, in particular
    - 2. Another recommendation was to hire a Chief Statistician: Dr. Karin Orvis, who comes from the Department of Defense, has experience in private sector, academia, and government developing evidence for policymaking
  - iv. Thank you to Bill Beach (ICSP) who will continue to be involved; Brian Moyer, acting duties for international work; Emilda, Chair of ACDEB
    - 1. Full confidence in Emilda's leadership of this Committee and have asked her to continue to lead

2. ACDEB has already done a huge amount of work, has drafted recommendations, on-track to complete Year 2 Report, OMB doesn't want to step in at the end and shift leadership
  3. About autonomy of Committee; ACDEB can make recommendations without OMB being present every day so that OMB can be recipient of recommendations with deliberate separation
- c. Remarks from Karin Orvis, Chief Statistician
- i. Thrilled to be joining OMB and very interested to learn from critical work of this Committee
    1. We all care about ensuring we have the best data to drive key decisions, whether those are decisions for federal agencies, state/local governments, private sector, or individuals in their own personal capacity
    2. At the same time, doing this work while ensuring we're protecting privacy, confidentiality, and strengthening public trust with respect to data
  - ii. Currently planning what CIPSEA 2018 implementation will look like in coming months and longer term
  - iii. Expand on Dom's point that would love to be part of this group for last several months, but it is more important to ensure ACDEB has full autonomy to bring forward recommendations for OMB's consideration in an independent manner
  - iv. ACDEB's recommendations will inform what next year and beyond is going to look like; looking for concrete and actionable recommendations because OMB/ICSP will be moving this work forward
    1. ACDEB is envisioning many aspects of a future data service; want to make sure the Committee is including things it would consider basic steppingstones
    2. For example, clear findings, specifics about state and local government data needs, clear recommendations on priorities for privacy-preserving methods, priorities for federal statistical system and its partners, goals and principles for regulations
    3. Even before we get to NSDS, recommendations will inform pilots, so actionable recommendations will help inform how to move from today's work on CIPSEA toward future we're envisioning together
- d. Thank You to Dom and Karin

- i. Very excited to have Karin as new Chief Statistician and look forward to advancing evolution of data ecosystem, goals of the Evidence Act, and federal statistical system together
- ii. Thank you to Dom for his leadership and delegation of Chief Statistician role across the statistical system, which enabled us to envision a more seamless system, to emphasize collaboration and partnership, and to tighten coordination
- e. Alyssa Holdren, ACDEB DFO, reviewed ground rules for member and public comments and questions

**II. Recommendations Summary with Committee Discussion** – Emilda Rivers, ACDEB Chair, ACDEB Co-Chairs

- a. Recommendations Summary
  - i. Today, will introduce recommendations with initial approval, will hear from subcommittee co-chairs on findings behind these recommendations, and will present big ideas around the governance and technical infrastructure implications for NSDS data capabilities
  - ii. Important to remember statements are drafts and additional editing will occur between now and October
- b. NSDS Organizational Structure Recommendation: “The America’s Datahub Consortium (ADC) should serve as the foundation for a future National Secure Data Service (NSDS) and should over time grow, adapt, and evolve to offer the capabilities and services necessary for it to realize the Advisory Committee’s vision for a NSDS.”
- c. NSDS Organizational Structure (Julia Lane, co-chair of Governance subcommittee)
  - i. In order to institute governance, have to have organizational structure
  - ii. Reviewed options that have historically been available and much interest in rethinking them
    - 1. Biggest concern was enabling different approaches to be used to empower innovation, empower success, and shut down failures
    - 2. Looked at international models, influential work of Nancy Potok, many other reports
  - iii. America’s DataHub is new approach being developed; nascent in nature and many details to work out but very promising. This model empowers government to bring in best and brightest from public, private, and nonprofit sectors to come up with new ideas; develop products that have value; and then shut it down if it fails

- iv. Dealing with a “many-headed beast”: want to develop something that can create value, try different access modalities including onsite access to different types of data, test out different measurement approaches and different linkage methods, and experiment with different privacy preserving technologies
  - 1. Building a system with multiple functionalities (networked hub) is what’s attractive about ADC approach
  - 2. Advantage is ADC can adapt, evolve, innovate, bring in best and brightest but not get stuck in monopoly or monolithic set of options
- d. Communications—Clear Purpose Statement
  - i. Thank you to Julia and will now looking at set of recommendations on communications
  - ii. First recommendation: “OMB should adopt a clear statement of purpose for the NSDS that is rooted in its core value proposition.”
- e. Communications—Clear Purpose Statement: Findings (Kim Murnieks and Matt Soldner, co-chairs of Other Services/Capacity-Building subcommittee)
  - i. Conducted extensive conversations with experts in fields of communications around data and data advocacy
  - ii. Concur with their judgment that OMB should take the initiative to clearly communicate about the purpose of the NSDS
  - iii. In Year 1 Report, describe NSDS as philosophy, place, and service; focusing on service is most powerful approach and way to lean into value proposition for federal, state, and local policymakers
  - iv. Stating purpose is foundational so it can inform work to come and clearly articulate to all stakeholder groups what the data service is about
  - v. From state perspective, like focusing on service because key to ensuring success is to show value and utilize service to inform state’s own public policy
- f. Communications—Public Presence: “OMB should create a public web presence for the NSDS that can serve as a hub for information about, and communications from, the Service.”
- g. Communications—Public Presence: Findings (Kim Murnieks and Matt Soldner)
  - i. Experts like those from Results for America and Data Quality Campaign emphasized need for a website (“NSDS.gov”)

- ii. Website would serve as a communications vehicle, as opportunity to educate stakeholders on role of NSDS in data ecosystem, and as landing spot for intelligent front door to help users easily access a range of services
  - iii. Having good web presence will help to manage the workload; can put frequent questions and information out there and lessen workload for staff
  - iv. Needs to be well-maintained so it continues to be relevant
- h. Communications—Comprehensive Strategy: “OMB should build a comprehensive communications strategy for the NSDS.”
- i. Communications—Comprehensive Strategy: Findings (Kim Murnieks and Matt Soldner)
- i. Received great feedback from experts on communications and data advocacy
  - ii. Not just communications strategy to help support development, energy, and resources to stand up a service but also communications strategy that spans entirety of its launch and then ongoing goals; includes everything from developing communications artifacts to identifying intermediaries and champions to spread the word of value proposition
  - iii. NSDS may start small, but it is a large endeavor with big hopes and potentially a real value to public, so needs to be matched with a communications architecture that is similarly robust
    - 1. Encourage OMB to resource communications function efficiently so service can be useful and used
    - 2. Need to make sure folks know about it and what it has to offer
  - iv. Understand that building a communications strategy and connecting with stakeholders is a big endeavor, but it is what enables a wider community to understand what NSDS is and isn’t from the beginning, which will lessen misconceptions; strong communications strategy at the start will pay off in the long term
- j. Communications—Stakeholder Engagement: “The NSDS must build a system for routinely engaging with stakeholder groups and user communities for the purpose of needs sensing, operational improvement, and advocacy for the use of data to improve policymaking.”
- k. Communications—Stakeholder Engagement: Findings (Kim Murnieks and Matt Soldner)
- i. Can’t underestimate importance of stakeholder engagement

1. Must connect with federal, state/local, and other stakeholders to ensure they are aware of service, to get feedback on what NSDS should or shouldn't be doing, to ensure NSDS meets needs
  2. Value must be communicated upfront, whether NSDS offers value must be monitored along the way, and must take feedback along the way
  3. Hesitated to use term "stakeholder engagement" because doesn't feel like quite enough; want to communicate even more strongly how important it is to be connected to groups engaged in this work
- ii. Communications doesn't just mean pushing out information, putting it on a website, stating purpose, goals, services; it is also listening, getting feedback, creating a two-way street that will inform long-term success of the endeavor
- I. Communications—Looking Ahead (Kim Murnieks and Matt Soldner)
- i. In next full Committee meeting, will discuss series of recommendations on technical assistance services
  - ii. Includes what user experience of NSDS should be like and ensuring all parties have equitable access so no one is shut out of benefits this service can provide
- m. Data Capabilities
- i. Thanks to Matt and Kim; will now look at recommendation around data hosting that is working its way through the Coordinating Committee and subcommittees
    1. A great test case for iterative engagement across ACDEB
    2. While wording isn't ready yet, process is driving toward tangible and actionable Year 2 engagements
    3. Will lay out considerations to help us calibrate these ideas
  - ii. Background: reminders of items from ACDEB's Year 1 Report and threads back to the Evidence Commission and Title III of the Evidence Act (CIPSEA 2018)
    1. In Year 1 Report, ACDEB stated NSDS should be a philosophy, service, and place within CIPSEA ecosystem
      - a. As a service, NSDS should have four functions: coordination, communication, research and development, and data standardization

- b. ACDEB envisioned NSDS would provide services to data users, data providers, and related communities of practice in support of evidence building
      - c. How the NSDS provides services comes down to the fundamental question of balancing tradeoff between utility and privacy
    2. Evidence Commission asserted NSDS should not be a data warehouse; important for ACDEB to explore deeper implications of that statement and answer questions:
      - a. Are there minimum viable requirements for NSDS to hold data?
      - b. What is required for validation and verification?
      - c. What is required for code transfer?
      - d. How to set the appropriate controls with governance and technical infrastructure?
    3. Title III of the Evidence Act (CIPSEA 2018) gave new authorities and responsibilities to statistical agencies; since NSDS will function within CIPSEA, ACDEB must describe how the NSDS will leverage authorities to deliver services
  - iii. Key Questions:
    1. What is the potential value of hosting data across the evidence ecosystem?
    2. What are the challenges?
    3. What are the possible solutions?
    4. What is the role for the NSDS?
  - n. Data Capabilities: Value and Risks
    - i. Big ideas around the potential value and challenges for data hosting (not comprehensive list, but types of discussions happening across ACDEB):
      1. Value:
        - a. In validation, verification, replication, so that if you do a study can you repeat those findings

- b. If thinking about data access, linking, and analysis services, especially for lower-capacity users, then what do you provide in support of that service?
2. Challenges:
- a. Hear a lot about privacy and confidentiality, and does that mean multiple copies of confidential data? Many concerns when posed that way, so thinking about what can we do to address that challenge
  - b. From cybersecurity point of view, do we want to have a single point of failure? These are big questions for us as Committee to address.
- o. Data Capabilities: Possible Solutions
- i. Key takeaway from these discussions is that there are both governance and technical solutions that can address these challenges; our task is to calibrate these solutions
  - ii. Governance:
    - 1. We see transparent governance models and policies, being clear about what we're doing, when, who is involved, what types of decisions we're supposed to be making
    - 2. When we look at statistical system, laws that govern data are different depending on Census data, IRS data, etc. so how do we remain consistent in governance with different types of laws
  - iii. Technical infrastructure:
    - 1. A lot of conversation about tiered access controls, what does that mean in terms of new paradigm of "yes, unless," how do we ensure access controls are implemented properly?
    - 2. Gets into conversation about privacy-preserving technologies; some colleagues have moved far with differential privacy, but there are other types of techniques as well; how much do we consider in implementation of those techniques in terms of balancing privacy and utility of data?
  - iv. Emphasize that as a CIPSEA entity some decisions will be handled by forthcoming Evidence Act Title III regulations and guidance
    - 1. For example, Access and Confidentiality regulation will lay out frameworks for data sensitivity levels, tiered access, risk assessments

2. Regulation is being drafted, so a great opportunity for ACDEB to advise on these items that will cover the whole CIPSEA environment, including NSDS

p. Next Steps

- i. May be thinking there's nothing new here:
  1. Always about privacy/utility tradeoff, Evidence Commission and Evidence Act spoke to this, so what's different now?
  2. Advancing "here and now" objective toward envisioned state is difficult; changes happen all around us and can make it feel like target keeps moving
- ii. Three advances that demonstrate what's different: work is complicated but also creates opportunities
  1. Federal statistical system is making great strides on implementing CIPSEA 2018; system is working more seamlessly and pieces of the data service are being built even before this new entity exists; forthcoming CIPSEA regulations will accelerate this progress
  2. There have been major advances in developing privacy-preserving technologies, and work in this field is evolving rapidly; technologies have the potential to offer new ways for users to access, link, and analyze sensitive data for evidence building
  3. There's a particular emphasis on the role of state and local stakeholders that was previously lacking; ACDEB has been instrumental in emphasizing this essential perspective

q. What impact can this Committee make to advance the discussion?

- i. While there's a heightened focus on evidence building right now, and we're seeing parallel conversations and content throughout the data ecosystem (for example, with CNSTAT, the CDO Council, and ICSP), there are different players at the ACDEB table and we have different perspectives to offer
- ii. While it seems like we're retracing steps of those who have gone before us, we've established a strong foundation from this earlier work and have broadened the scope of the conversation
  1. For example, highlighting the role of state and local governments; let's use ACDEB's voice to be clear about who is necessary in these discussions.

2. Statistical agencies have been coordinating with states for decades, but where are the gaps in stakeholder engagement? Are the right state and local people involved in making key decisions?

- iii. Provide a nuanced, robust set of recommendations; not about the governance piece or the technical piece but about leveraging tools at our disposal to offer a portfolio of solutions that is flexible enough to stand the test of time
- iv. Year 2 Report is opportunity to offer formal input into broader conversation; help bring the pieces together so that those who come after us have a starting point that is miles down the road from where we began our journey

r. **Committee Feedback/Discussion**

- i. **Comment:** on communications recommendation, suggest adding to the vision statement about the value of research (broadly defined—within and across agencies)

On access, convey that these are resources for use under certain requirements, but make it clear what resources are available and how to use them

On Evidence Commission, we live in a world with multiple points of access to confidential data, and there are spillover effects; can gain greater protection by limiting access points—important to think about when considering technologies and where data are accessed

**Response:** agreement among committee members in chat regarding research in the vision statement, communicating how people can get resources

- ii. **Comment:** agreement among members in chat regarding comment about communication being a two-way street between state/local and NSDS
- iii. **Comment:** on governance, the Governance subcommittee will be unpacking different ways value can be generated; list on slide was illustrative not exhaustive

- s. **Wrap-up:** Many thanks to presenters and comments we received; will have more conversation around these points as we refine recommendations

III. **Subcommittee Report: Governance, Transparency, and Accountability** – Charles Cutshall, ACDEB Member, and Julia Lane, ACDEB Member

a. Overarching Findings

- i. Function: providing access to data for evidence building and programmatic improvement for federal, state, and local governments; broad focus is

evidenced by composition of this Committee, which includes federal, statistical, and programmatic agencies, as well as state and local governments

- ii. Value: driving idea is that NSDS should be front and center providing value produced by evidence
  - 1. Use cases: wanted to get concrete, so working on three use cases in conjunction with other subcommittees: health, labor markets, education/workforce (represent competencies in subcommittee)
  - 2. Listed value proposition and will be unpacking that to think about, how do we bring together data from variety of sources and use it in robust, sensible, and reproducible way to improve research, evaluation, and programmatic delivery
  - 3. One big advantage is reducing resource and cost burden to public because focus isn't just on research and statistical data but also on programmatic improvement
  - 4. One issue NSDS needs to address is minimizing delays to accessing data (for example, needing information in weeks not years during pandemic), so how to build infrastructure to do that
  - 5. Ever-present has to be balance between privacy/confidentiality and value; a joint discussion
- iii. Requirements: transparency and trust, scalable, sustainable, strong oversight and accountability, intergovernmental support, legal authority to protect privacy and confidentiality, accessible, innovative
- iv. Disincentives: resource constraints, legal barriers, considering private and public sector incentives related to bureaucracy, and political issues—all core challenges organizational infrastructure should address

b. Recommendation: Governance

- i. America's DataHub Consortium will need centralized management structure
  - 1. ADC does not address that as it currently stands
  - 2. There is a contracting mechanism but not a managing structure, so will spend rest of time between now and July discussing what centralized management structure should look like, what governance body/process should look like, and structure to coordinate nodes
- ii. Governance structure should be set up to continuously produce value; as we've seen, the opportunity to produce value has been vastly expanded, and ADC would have to continuously be monitored to produce that value;

commissioning additional nodes might be a very sensible approach to doing that

iii. To Dom's comment and President's budget, this won't exist without federal appropriations both to federal statistical agencies, programmatic agencies, and state/local governments

1. The way in which that should be structured should be a point of discussion
2. Agencies at all levels have been starved of resources, so it is very important to ensure that they are not starved moving forward through right incentive structure, not just reliance on federal appropriations

c. Proposed Governance Model: Central Management Structure

- i. Preliminary suggestions and would like feedback from Committee and public
- ii. NCSES should be provided resources to set up project manage office, policy steering committee
- iii. Contractor entity would manage day to day of individual hubs, under Board of Directors and Research and Technical Advisory Board to address rapid technological growth
- iv. Community engagement is core and central to operations
- v. Projects should be approved by Committee and include appeals process under Project Appeals Board
- vi. Stakeholder operational input structure: many stakeholders to consider; NSDS is "national" so lots of stakeholders, and is a "service" so lots of communities

d. **Committee Feedback/Discussion**

- i. **Comment (from chat):** agreement with comment that there must be resources to build capacity at all levels of government
- ii. **Question:** can you say more about Project Approval Committee; would it be for all uses of NSDS, uses in a certain class, or certain level of resource allocation? How would that function work?

**Answer:** NSDS can't be all things to all people out of the gate, so will have to be a few initial projects that are identified and resources provided; stakeholders would have to identify a few high-value projects and then as community has ideas, they would be submitted to approval board and that approval process would go through whatever governance system is set up

- iii. **Comment:** thinking about idea of NSDS as federation of multiple nodes complicates that a bit; could envision a user being re-routed to a different group to meet their needs; as you think through it, it will be interesting to see the governance around this

**Answer:** will be a challenge; if using state data, states have priorities and things must fit within their legal framework, so approval process would have to go through state approval as well

Shawn Davis is looking at Midwest Collaborative approach (Anna Hui is lead on that group); have very promising infrastructure we expect to be drawing on going forward; at least have a starting entity to consider these possibilities

- iv. **Question:** like proposed structure; as more general question, how much detail does ACDEB need to provide? Outline general rules about governance structure, but ADC is very flexible (good) so not sure how detailed we should be in prescribing governance structure of each entity; how much more structure does ACDEB need to provide beyond general principles?

**Answer:** has to be a centralized management structure; don't need governance for individual entities (FFRDC, nonprofit, government agency, might each have their own governance structure, for example), but need a central governance structure to manage operation of these individual nodes and to ensure a core adherence to the values and principles of NSDS

- v. **Question:** agree; what beyond that is next/more for governance?

**Answer:** don't want to be overly prescriptive, but at the end of the day, organizations live and die by quality of their governance structure (definition of mission, establishment of KPIs, explicit engagement with stakeholder groups), so that must be baked in from the beginning

There are different views on this (unwritten vs. written constitutions, for example, with plusses and minuses for both); should put down defining principles in Report

Just as we're paying a lot of attention to principles of the Evidence Commission; want ACDEB principles to be as influential and as high quality

- vi. **Comment:** agree with set of principles and need to move beyond Evidence Commission report, but how successful ACDEB will be is a function of not being too prescriptive; struggle with finding the balance

**Response:** completely agree

- e. **Wrap-up:** Many thanks to those providing comments and will pass along public comments; will continue thinking about how we move forward and balance we need to strike

**IV. Subcommittee Report: Other Services/Capacity-Building Opportunities with Committee Discussion** – Kim Murnieks, ACDEB Member, and Matthew Soldner, ACDEB Member

*Note: Previewed where we're heading in earlier presentation, so yielding time*

**V. Subcommittee Report: Technical Infrastructure** – Amy O'Hara, ACDEB Member, and David Park, ACDEB Member

a. Status Update

i. Our subcommittee has heard from many experts on ways NSDS would need to build a secure service that does lots of things:

1. Datavant COVID-19 Research Database: stood up within a number of weeks at beginning of pandemic; uses highly sensitive private sector data and enables researchers to use data in secure virtual setting that has generated hundreds of insights; quickly and nimbly used PPTs and more traditional practices (don't have a fully automated system)
2. Advanced Educational Research and Development Fund (AERDF): improving how to deliver evidence in education field; good contrast on current state and the future by implementing a number of PPTs to help protect underlying data and prevent disclosure on what is published

ii. Spoke with experts on quantifying privacy risks:

1. Urban Institute, folks at Census: heard about how we can identify privacy risks in what has already gone out the door by doing reidentification studies
2. How to do privacy risk assessments; important when thinking about protecting data you're using and publications that are generated are not going to allow for reidentification

iii. Spoke with experts on synthetic data and validations servers:

1. Urban Institute project at IRS and work at Census microdata: a lot of great work going on with synthetic data; tons of governance issues there, but all the folks we heard from made it clear that when you have synthetic data it will be important to have it in parallel with validation servers that allow you to test if queries on synthetic data hold true on underlying data
  - a. When running on actual data, results still need privacy protection
  - b. Group has been looking at how different technologies need to click together, depending on use and need for accuracy; ties

into conversations and learnings about PPTs more generally, which was our use case from year 1 report

- iv. Spoke to number of different groups looking into shortlist of PPTs: What is viable today, what is current state of R&D to understand how they are improving to be viable and scalable solutions, and know there are projects underway across federal statistical system
- v. Have more expert visits planned to hear even more information that can help our group write up what we believe technical infrastructure principles and priorities can be to align with Five Safes framing all subcommittees are using

b. Findings

- i. NSDS needs approaches to conduct linkages across different data types
  - 1. Have heard from experts who have this occurring at scale in highly secure ways
  - 2. PPRL as demonstrated by COVID-19 Research Database
  - 3. Want to make sure methods of doing linkages across data types are affordable and have transparent process so that end users understand any errors introduced in linkage process and how to mitigate those
- ii. NSDS needs tools that guide decisions about which privacy-preserving solutions are suitable
  - 1. Likely to include discussion of privacy risk assessments; within those, what sort of disclosure risks are you trying to handle (consideration of identity, attribute, and inferential disclosure risks)
  - 2. Should avoid overkill solutions and ad hoc decisions/privacy parameter settings (how to set privacy budget, what is the cell size approach; are there ways to have standards around that)
- iii. NSDS needs to move beyond current practice
  - 1. Some experts discussed privacy-aware data synthesizers, which automate process of creating synthetic data, baking in privacy at the front end
  - 2. Heard about how to design sample surveys with privacy baked into that sampling; all research right now and not application but goes to show how fast this is all moving
  - 3. Also heard about diagnostics, checking data itself or checking diagnostics on how privacy-preserving setup is functioning

4. Incredible opportunity for cross-discipline work and knowledge sharing; has been siloed in the past, happens with PPTs and data linkages—so how can we improve that community of practice and get the good and the bad to be put on the table (what works, what doesn't) and all move that forward, especially exploiting deep investments that already exist in federal statistical system

c. Next Steps

- i. Continuing research on use cases (PPTs)
- ii. Upcoming meetings include NIH N3C and Social Science One (privacy protecting data access project involving social media data)

d. Discussion Questions

- i. Does anyone want an overview of what PPTs are? Intend to write a memo to ACDEB to get this in writing, but if any subcommittees would like a short briefing in their subcommittee meetings, let us know
- ii. What about other groups' use cases; discussed that Technical Infrastructure could look at use cases and propose how PPTs could be inserted in the future
- iii. Have been thinking about Five Safes and FAIR data; where is the Committee on FAIR issues?
  1. Curious when we think about findable data are other subcommittees working on data discovery? If we need to be weighing in or exploring the existing inventories and how technology might be able to aid discovery across them, that would be something to talk about
  2. Accessibility: ties in with technical capacity and user interface; should some offers of technical assistance be tied more closely to services themselves? If there is a need to do connection across subcommittees there, let us know.
  3. Curious about way that tiered access could be framed in the Year 2 Report; want to make sure when states/locals need query results (quick number on a regular basis), how could NSDS make sure one of the tiers of access supports these quick and cheap/free (to users) query results?
  4. Thinking about interoperability: when thinking about overlap between programs, will NSDS be able to run complex statistical programs at scale? Not just for a sample but on populations; thinking about technical infrastructure there, which ties into safe people, safe settings, safe outputs (what are principles that need to be put in play for that)

5. When you have a validation server for synthetic data, that validation server needs to hit real data on the back end; curious whether that is going to be NSDS facilitating or NSDS is place for those data to reside? Validation ties to “reusable”—what does this mean to other subgroups?

**e. Committee Feedback/Discussion**

- i. **Question:** very useful presentation; much of what Technical Infrastructure is doing is assessing cost to using data and keeping it private (important); there are costs and benefits, so how do we capture those benefits when thinking about this inherent tradeoff; coming up with better and better technologies, but how to assess the balance?

**Answer:** if you’re going to build that structure to catalogue “wins,” need to have real community engagement to catalogue “harms” (beyond cybersecurity threats)

Need to think about each project weighing privacy/utility risks; there are already harms happening now (equity, for example); studies may root out fraud and abuse but can be harming communities; need to make sure value has a plus and minus attribute attached to it

- ii. **Comment:** agree; need places in recommendations that emphasize here’s what we’re getting by creating this system

**Answer:** something we grapple with because that’s part of the technical infrastructure in as much as NSDS needs to build searchable index of such things, have standards for how such things are put into repository, but don’t know which other subcommittee has pen for the “what” and “why”; Technical Infrastructure can talk about nitty gritty on “how”

- iii. **Comment (from chat):** sent the coordinating committee the BRDI report on automated reusable workflows which addresses a lot of these questions, particularly focusing on FAIR issues.

**Response:** report is definitely valuable, and questions from Technical Infrastructure still need to be grappled with in terms of how they fit into the existing ecosystem and legal framework

**f. Wrap-up**

- i. Thank you for highlighting topics important to recommendations from ACDEB to OMB/ICSP; thank you for recognizing many interdependencies involved in conversation

- ii. This is challenging work, and everything we put forward does not have to be completely worked out but can include nods toward those who will carry this work forward (OMB, ICSP, CDOs, state/local, etc.)
- iii. This is the type of thinking that leads to actionable recommendations, which will look like “here and now” as well as pointing to the future
- iv. These conversations are important to have from the beginning and discussion of benefits and potential harms needs to be clear

**VI. Subcommittee Report: Government Data for Evidence Building** – Anna Hui, ACDEB Member, and Kenneth Troske, ACDEB Member

- a. Status Update: subcommittee has met with four experts over the last several weeks:
  - i. JEDx
    - 1. Potential public/private partnership to think about value proposition and challenges of those types of partnerships
    - 2. U.S. Chamber of Commerce to standardize employment information businesses are required to report
    - 3. Trying to minimize number of times states have to put their data in; instead, submit to centralized portal once and it will report to agencies
    - 4. Big value-add for businesses because minimizing number of times they must report
    - 5. Payroll processing firms would do the data collection and reporting, which has some issues as well
    - 6. Challenges:
      - a. Raises issues of standardization: requires agreement around what a “job” is, what an “occupation” is, but leads to potentially better, higher frequency data using standard definitions
      - b. Part of problem is federal and state agencies have different statutory requirements, so can we standardize what a “job” is across these spaces?
      - c. Other challenge is this is a process that works well for large businesses but might not work well for smaller businesses that don’t use payroll processing

- d. Coming from the private sector, so difficult to get buy-in from public sector
  7. If we could get these data, we'd have information for over 90 percent of workforce but only 50 percent of businesses
  8. Started a promising pilot with seven states to test it out, focusing initially on unemployment insurance data
- ii. FY 23 Budget
1. As we heard already, the use of evidence-based policymaking is rife throughout various parts of the budget, which is great
  2. Most of the discussion related to ACDEB/NSDS is in chapter "Leveraging Federal Statistics"; part of what is important is it emphasizes re-establishing trust in federal statistical agencies
  3. Major budgetary impacts include:
    - a. Investment in NCHS to develop virtual data enclave
    - b. Additional money for SOI at IRS to hire more staff, which would help provide more resources to expand researcher access
    - c. Increased funding for NCSES for NSDS demo and to support ADC
- iii. SWIS
1. A version of an NSDS in some ways; an agreement among all 50 states to share UI data for reporting outcomes for WIOA
  2. Organized, facilitated, coordinated between Department of Labor and Education but is really an agreement among the states to sign and implement
  3. Using centralized ICON system to identify movement of workers
    - a. Know which states have UI data for an SSN and then match SSNs across state lines to track outcomes
    - b. Doing record linkages at different locations; a centralized component but then actual information is provided by individual states for program administration (reporting success of WIOA participants)
  4. Has components of what NSDS might look like with a bunch of different nodes

- a. Not designed for research or program evaluation; that is possible but a lot of hoops to jump through before you could do that
    - b. But does demonstrate that it's possible to get states to agree to share data (agreements are often for a very specific purpose and for very limited use)
  5. Presumably all states could do this
    - a. That's the type of thing that Midwest Collaborative is trying to do through an alternative method
    - b. Hits on value of communication and coordination
- iv. PFAS Project
  1. Environmental and health outcomes data from EPA; how it would potentially benefit from an NSDS; federal trying to get data from states
  2. PFAS substances are ubiquitous in the environment and presumably harmful; "forever" chemicals because they don't decompose
  3. There are known and expected toxicity issues that impact human health, but not enough research on human health impacts and how PFAS get into the system in a variety of ways; for example, manufacturing processes or through consumer products (GORE-TEX, pizza boxes)
  4. EPA wants to document this, but they need to link multiple data sources (environmental and human health); there is value for role of NSDS to understand the impacts of PFAS
  5. Value and challenges:
    - a. A lot of data held at the state level
    - b. Some states willing to share, other states less willing to share
    - c. Better linkages across federal agencies (CDC/EPA/Census) and better linkages between federal and state
  6. EPA is part of ICSP efforts (mandated by Evidence Act); if allowing non-CIPSEA agencies to be part of process around SAP, could we open process for state and local governments as well?
- b. Discussion Questions

- i. What things can we do to improve access to data? Start thinking about this problem more holistically; what are net benefits (gain access here by giving access there) to sharing data?
- ii. Ways to think about better cooperation/collaboration/willingness to share?
- iii. How to encourage parties to more widely adopt mindsets that promote getting to “yes” versus tendency to start with “no”?
- iv. How could NSDS help balance what is mandated by statutes that are doing things that are not excluded?
- v. Final thoughts: a lot of examples are great and moving into a pilot stage; ongoing concern is that resourcing and capacity-building among partners isn’t always equal, so need to distribute resources, think about investment in parity and equity, especially in vertical chain (federal and states)

**c. Committee Feedback/Discussion**

- i. **Question:** if there was low or limited capacity at state level, but NSDS could do “things,” would states say “here’s our stuff, do things for us”? Would they do this on an as-needed basis? Or does this skip investment at the state level and would states rely on NSDS? Can spin ADC up or down, kind of an experiment, so thinking about do states want to be dependent on something potentially fragile?

**Answer:** Federal/state partnerships need active engagement on both sides; have needs at state/local levels that must be met, and NSDS may not have capacity to service all of us, so it’s about finding the right balance on both sides and technical assistance that can be provided, knowing there are bigger picture questions that could be coordinated through NSDS

Ongoing issue is that if you want good quality data collected and reported “up” so federal statistical system can use it, can’t neglect investments at state/local level (infrastructure, technology, and human capital)

Need ongoing support from federal level to states because states will have needs beyond those handled through NSDS, and also about considering how states can meet some immediate needs on their own (e.g., leveraging American Rescue Plan money to build technical infrastructure)

About partnership/collaboration, about addressing gaps, and about communications piece of ongoing dialogue/feedback

- d. **Wrap-up:** Thank you for raising these important issues; want to comment on a few things:

- i. Mention of EPA as a non-federal stat agency; thrilled EPA is represented with statistical official and has a seat on ICSP; a great opportunity to address issues raised here
- ii. Hearing issues of state/local capacity throughout these conversations; federal statistical system must collect data from state/local governments and needs to be a two-way conversation, so looking for ways to engage around that; think about specific groups with which to engage
- iii. Idea of what is the balance we need when thinking about what services NSDS would provide versus services provided by other actors in the data ecosystem
- iv. Also note role of ONET mentioned in the comments; familiar with the data—part of what the community looks at in terms of resources (ties into SWIS)

**VII. Subcommittee Report: Legislation and Regulations** – Nicholas Hart, ACDEB Member, and Christine Heflin, ACDEB Member

a. Status Update

- i. Have revisited a topic that has been under discussion for decades: data synchronization, or data harmonization
  - 1. Included in FY 23 Revenue Proposal (Green Book, pg. 78)
  - 2. BEA/BLS limited access (BEA only for corporations and BLS not at all) creates challenges around the accuracy of economic statistics; more and more businesses aren't corporations, so this effects outcomes as well
- ii. Availability of administrative/statistical data for evaluation:
  - 1. Large effort at Commerce; Department of Commerce task force to look at potential for using statistical and administrative data to evaluate impacts of Infrastructure (NTIA) and American Rescue Plan (EDA)
  - 2. Hardly seamless, so talking about what could one do: what would need to change? Would need to change laws? Additional guidance from OMB? Technologies to enable that process? Leveraging policies like special sworn status (who gets SSS and under what circumstances)?
- iii. Appropriations to support data management:
  - 1. All agencies (not just statistical) that have major datasets, need funds to structure and manage data

2. Upfront investment will save future costs in terms of time and energy to use data for any number of purposes
  3. Have Evidence Act but not everything in place to make it work as intended
- iv. Federal/state data for evaluation:
1. Opportunities for collaborating with states and using data for evaluation are exciting
  2. Very local effects; can't look at county data, very localized, most current and specific data is at state level
  3. Shared stake in knowing what works; NSDS could facilitate discussion with states on laws/rules that would support collaboration
- b. Findings
- i. Sharing business tax information among Census/BEA/BLS—"data synchronization"
    1. Three agencies would be subject to taxpayer privacy laws and CIPSEA safeguards, requirements, and penalties
    2. More people would have these responsibilities but added value in terms of statistical accuracy may be worth it
  - ii. Availability of administrative/statistical data for evaluation
    1. For the most part, statistical information is available, but we would have to jump hoops to use data, so how to make this more readily available?
    2. Don't want to impose reporting burdens on recipients, especially as we become more inclusive with communities with fewer resources, so again what can we do to make that information more readily available?
    3. Possibilities: Special Sworn Status might help, technical facilitation, some of it might be eased by direct instruction from OMB giving agencies more comfort, bureaucracies have conservative nature, or might need to change law
  - iii. Appropriations
    1. Potential downstream savings and ROI—pay now or pay more later
    2. Potential of data for competitive advantage is enormous but need to invest upfront in data management

iv. Federal/state data for evaluation

1. What role could NSDS play in more collaboration and better collaboration with states?
2. Could build models with different alternative laws and rules depending on state's interests

c. Next Steps

- i. Will be listening carefully and having further discussions with ICSP on guidance regarding data sharing for evidence and evaluation
- ii. Discussing with experts at Commerce on Special Sworn Status, laws and policies that govern that; Harvard project with in-depth information on those laws and rules
- iii. Would like more discussion with ACDEB state representatives about barriers to and advantages of collaboration on program evaluation; what works and what the challenges are in different circumstances and locations
- iv. Documenting use case; Commerce report on using statistical and administrative data for evaluation (in clearance now)
- v. Recently learned about Curated Data Enterprise; involves using more datasets in development of US statistics; what are the implications of that for evidence building?

d. Discussion Questions

- i. Data synchronization: Is this the right path? Does it go far enough? What concerns do we have about it?
- ii. How directive should OMB/ICSP be in their guidance on data sharing? How much discretion should be left to agencies knowing there's conservative bias there; how hard a nudge should OMB give?
- iii. How can use of administrative/statistical data for federal evidence/evaluation be facilitated?
- iv. How can federal/state collaborations on program evaluation be facilitated? How to address individual states' reservations or legal/regulatory constraints?

e. **Committee Feedback/Discussion**

- i. **Question:** any thoughts from Technical Infrastructure on data synchronization?

**Answer:** data sync has been a longstanding request; hope this receives favorable review and achieves this data sharing

**Response:** ACDEB could help give an extra push

- ii. **Comment:** on data sync, believe there is draft legislation by Congressman Byer; recommendation from Treasury to do this, don't know if it's been introduced; may be looking for co-sponsors

Would like to see synchronization go even further: why should Census get data that BEA/BLS don't? Important for maintaining one business list versus two; could use money we save to provide more resources for evidence-building

- iii. **Comment:** believe it also has to do with sampling, different sampling frames

**Answer:** recent paper on Payroll Protection Program: main comment was about difficulties with not having a common business list; \$200K per job saved, so clear benefits from having data synchronization that would have improved implementation of this program

f. **Wrap-Up**

- i. Want to recognize there is a lot of work happening within the federal statistical system on data synchronization and related topics (Standard Application Process, a lot of other laws that come into play beyond Special Sworn Status); need a holistic approach with paradigm of "yes unless," so thank you for starting that conversation
- ii. Thank you for flagging that Evidence Act provided framework for agencies with Statistical Official, Evaluation Officer, and Chief Data Officer to begin to move toward "yes unless" with making more data available, but a lot more work to do and that means resources and capacity; thinking about the "here and now" and the future state of evidence making; thinking about those building blocks you've started to raise here
- iii. Idea of respondent burden is a big one that gets back to conversation around collaboration; the more collaboration, the better data collection; one big aspect around sampling is sample frame and who has access to sample frame
- iv. **Question (from chat):** were there a lot of responses to the Standard Application Process Federal Register Notice? Will ACDEB be able to see a summary of those before the next meeting?

**Answer:** will check on this and can likely have summary of whatever is available before next meeting

**Answer:** Alex Marten is working on this now; quite a few responses and working to summarize those into buckets and update draft policy document; seems doable before July meeting

**VIII. Piecing It Together: Scenarios – Emilda Rivers, ACDEB Chair**

a. The Evolving Evidence Ecosystem

- i. Reminders we've touched on throughout this meeting and placing ACDEB's work in the broader context of the evolving data ecosystem:
  1. Evidence evolution and the work to support it keeps moving
  2. As you can see from the "Items of Interest" in the Weekly Update emails and the Resource Library posted to ACDEB's website, many groups across the public and private sectors are focused on evidence building right now and there are many questions that remain; for example, ICSP is thinking intensely about the implications of the evolving roles for federal statistical agencies and for the NSDS, not to mention how to leverage upcoming regulations to advance shared goals
  3. Therefore, want to assure you that there is plenty of room for your best thinking on issues for today and beyond today
  4. This work will continue for a long time, so it's okay to tee up issues that this group can't fully solve right now; when the Committee disbands, the Interagency Council on Statistical Policy (ICSP) and others will continue to be responsible for figuring this out
  5. ACDEB shouldn't underestimate the value of findings and recommendations that may seem obvious to you as experts in particular fields, as well as the value of raising issues for those of us who will continue this work moving forward

b. Scenarios: Tool for Creating a Shared Vision

- i. Want to highlight a tool that can help us develop a shared vision for the NSDS and map out next steps for the data ecosystem more broadly:
  1. ACDEB subcommittees have been engaging in rich conversations around case studies and personas
  2. Scenarios weave these threads together by matching a user with specific needs to a topical area
- ii. Scenarios are valuable for several purposes, including:

1. Helping us evaluate and describe more explicitly how the NSDS fits into the evolving ecosystem and how it aligns with the CIPSEA legal framework
  2. Helping us better consider different users' needs, how the ecosystem functions now to address them, and what this could look like in the future. (In other words, who does what between the NSDS and existing actors in the ecosystem?)
  - iii. Scenarios are also a tool for helping different stakeholder groups (like CDOs and ICSP) think through these same questions
- c. Scenarios: Tool for Creating a Shared Vision
- i. ICSP members have begun to discuss some of these scenarios with great insights already
  - ii. Even with "simple" scenarios like previous example, there are nuances and layers of consideration
    1. For example, on the surface it may seem like the state health official would reach out to the National Center for Health Statistics, and NCHS staff would meet the users' needs with publicly available data resources and in-house expertise (something statistical agencies do every day)
    2. However, it isn't that simple, as the vaccination data "live" at the state level, and there are multiple "death rates" to pick from
  - iii. Scenarios scratch beneath the surface of a user's needs
    1. For example, a core need from the state health official example is about data discovery: users must figure out what data are available to answer questions
    2. While there is no single way that this discovery process happens today (and probably no single way in the future either), a lot of folks start with a simple internet search
    3. Statistical agency leaders recalled prior efforts to work with major search engines to ensure high-quality, objective datasets had properly formatted metadata to show up at the top of search results
    4. This need as part of a full-service data and evidence ecosystem had not been front and center before using scenarios as a tool
- d. Scenarios: Fitting the Pieces Together

- i. How do these scenarios fit in with ongoing activities, both for ACDEB and other groups?
- ii. The support team will share scenarios with ACDEB’s subcommittees to bridge discussions around use cases, personas, findings, and recommendations the groups are building right now; this will help the Year 2 Report be more cohesive as well
- iii. Want to emphasize perspectives of all Committee members on the scenarios are invaluable for helping us envision the future
  - 1. There is a lot that will be figured out after this Committee sunsets, as the federal statistical system pilots, demonstrates, and applies ACDEB’s ideas
  - 2. Timely, relevant, and actionable findings and recommendations are what will help the ecosystem move forward over the coming months and years
- iv. Scenarios will also help tie together conversations across the various data communities that are going to be critical contributors for the future

**IX. Next Steps and Action Items with Committee Discussion – Emilda Rivers, ACDEB Chair**

- a. Our subcommittees are currently gathering information, developing findings, and drafting recommendations
- b. Once these recommendations have been approved within the subcommittees, each group will circulate findings and recommendations with all members through the Coordinating Committee
- c. This will help us refine the recommendations, sort out which ones are ready for full Committee approval, and identify which ones require more facilitated discussion
- d. July Meeting and Looking Forward:
  - i. By the July meeting, our goal is to have all recommendations submitted, reviewed, and approved by the full Committee to the fullest extent possible
  - ii. No new recommendations will be presented at the July meeting; at the latest, they should be provided to the Coordinating Committee in advance of their July 11 meeting
  - iii. We’ve got a lot of work to do over the next two months, but after July we’ll be on track to address lingering issues and package everything into a timely, actionable, and relevant Year 2 Report
- e. Next Meeting: ACDEB Meeting 18: July, 22, 2022