2025 Data Integrity Summit

The federal statistical system has been struggling with budget constraints for several years. In the first six months of 2025 threats to the integrity of the system increased substantially, including cuts to staffing and contracts, data removals, revisions to data collections, and delays and missed deadlines in the publication of key information. Day 1 of the Data Integrity Summit brought together data experts and users across the U.S. to strategize about recovering from the current disruptions and rebuilding an efficient federal data infrastructure for the years ahead. Day 2 took a deeper dive into issues of transparency and public trust.

Recurring Themes

Across the two days, four key themes emerged.

Transparency and public trust are paramount

Across all breakout sessions, transparency and public trust were identified as critical considerations for any data system changes. Clear communication (from planning through publication), broad public access to both data and documentation, and robust protections against misuse are essential to maintain public confidence.

Data quality is key

A recurring theme was the challenge of tracking and ensuring data quality in new or consolidated systems. This involves defining new metrics, conducting robust testing (e.g., parallel runs for surveys, comparing administrative data to traditional methods), providing error metrics, and publishing clear documentation on data methods and limitations.

Education is needed

While participants were all deeply knowledgeable about threats to the federal statistical system, many data users are a step (or more) removed from knowledge about data infrastructure and the current threats to data access and quality. More education is needed at several levels including among elected officials, membership organizations, and data users.

Privacy protection is critical

Given the increasing use of linked survey and administrative data, as well as changes in the way the government is using data, many participants called for enhanced privacy laws, strict data use agreements, and robust oversight mechanisms. This includes ethical/privacy review boards, community-informed oversight, and increased penalties for breaches to protect against misuse by government agencies or third parties.

Overall Action Items and Next Steps

There are specific actions identified for each discussion above, with themes that recur throughout. APDU's key action items fall into categories that mirror the key themes above.

Transparency and Public Trust

Short-term Actions

- Identify and collaborate with business ambassador(s) to help advocate for data for businesses and for workforce development.
- Develop a "roadshow" or "plug-and-play" presentation that anyone could give that can be tailored to local governments, philanthropic organizations, industry groups / associations, and others.
- Define transparency in a way that data producers, data consumers, and the general public can understand.
- Consider using the advisory committee model as a mechanism for keeping up pressure for transparency in decision-making.
- Track instances where regulations are not being followed and put pressure on OMB.
- Facilitate introductions between recently-released federal staff and journalists by making introductions, hosting mixers, and/or providing media training to former fed staff.
- Connect with Congresspeople who are engaged on transparency issues now.
- Include transparency and public trust in education and outreach activities (described below).

Long-term Actions

- Push for transparency at every step of the data lifecycle–from planning through publication.
- Catalogue and start to work through the hard questions:
    - How do we define transparency?
    - What do we do when transparency goes bad?
    - What do we do when there is malicious compliance?
    - When is transparency not enough?
    - When is transparency not the goal?
    - How should we be crafting transparency strategies?
    - What should the standards be for transparency?
- Propose a definition of and standards for data transparency, broadly. And also…
    - Propose standards for transparency around linked data, such as (but not limited to) minimum reporting standards.
    - Propose standards for transparency about the implementation of AI/ML tools. Most of these are "black box," so there needs to be robust processes (like making training data available) to ensure transparency.
    - Under this umbrella, there is also a need to support additional research into the implications of ongoing changes such as the use of AI and ML, use of

administrative data, consolidation across surveys, implementation of privacy-enhancing technologies, and the associated and implications of changes.
- Support the "Stat Influencer" idea for Tik Tok to influence younger generation.

Data Quality

- Ensure that any changes to data collection are based on a scientifically rigorous implementation strategy (and technical plan to permit benchmarking, quality assessment).
- For AI/ML learn more about what has already been done within the federal statistical system and what are the lessons learned that could inform future ethical use.
- Also identify the goal(s) of any changes–are the goals to reduce respondent burden? Reduce cost? Improve accuracy? Improve timeliness or granularity? Improve privacy? Be prepared to demand that those goals are clear at the outset and that there is evidence that the change(s) will support the desired outcome(s) *before* implementation.

Education and Outreach

Develop a strategic communications plan, that may include:

- Segment the strategy by audience.
    - Key audiences include: businesses, foundations, state/local governments, and academics.
- Include a risk assessment (will communications help or hinder progress).
- Hold bilateral meetings with left, right and center think tanks to find common ground.
- Conduct research into what messaging strategies are effective in this environment, and if we find nothing, hold focus groups to learn more. Subsequently, conduct training sessions for the media and data stakeholders on best practices.
- Update secondary school and college curriculum standards to make data literacy a core requirement.

Specific actions may include:

- Hold a convening and invite agency staff so we can learn what they need and want. Also to connect federal agency staff with data users so we can turn data users into data advocates.
- Compile primary uses for key statistics/data elements so that we can reference these as we develop messaging; flesh out use cases for specific audiences. (*Editor's note: Work of this type is underway–APDU's role may be to support and uplift those efforts.*)
- "Take APDU out to the streets," not just a yearly DC conference.
- Continue to host events where data users can learn about the latest state of affairs, share information, and strategize.
- Conduct outreach to partner and peer organizations (like NASBO, NCSL, and others) to raise their awareness of ongoing changes in the federal statistical system.

- Develop an open-access slide deck that anyone can use to share the latest on what's going on with public data.
  - COMPLETE - You can access the slide deck here: https://docs.google.com/presentation/d/1FgfUdhfXvJsWdoxHdeQeVJURo2jVqKYvfaTjNu9wLi8/edit?usp=sharing

Privacy protection

- Outline a plan for how enhanced privacy standards and oversight would function (work with Representative Trahan's office).
- Support efforts from community-led organizations to make sure all voices are heard.

2025 Data Integrity Summit - Day 2 Summary

When reviewing findings from Day 1 of the Data Integrity Summit, organizers found that the transparency discussion had the fewest concrete actions identified. For that reason, on Day 2 we brought together data experts from across the U.S. to strategize more specifically about transparency and public trust.

Transparency was not always robust coming into 2025. For example, some of the decisions related to the 2020 Census left the data user community feeling like there was a lack of transparency, despite efforts to make the process more transparent. In the administrative records world there were often times when data structures would change with no notice and little or no documentation. But there were also transparency wins–public notification of potential changes to disability questions in the ACS led to community mobilization that ultimately forced the Census Bureau to change course.

And on top of an imperfect system, 2025 ushered in new challenges. At the most bureaucratic level, changes in data collections being are being labeled as "nonsubstantive" when they are clearly substantive. DOGE tweeted that they canceled several data collections, but it's unclear which ones (or if the cancellations were averted). The Pregnancy Risk Assessment Monitoring System (PRAMS) data collection was suspended–which we only know through state partners, not because of communication from CDC–and there is no clarity on whether or not it has restarted. We know that record linkage for immigration enforcement is moving forward–not because there has been appropriate communication but because agency staff have leaked information to the press.

But even more fundamental than that, we often don't know what agency staff (or even divisions) are still standing. FOIA requests–which are the most powerful tool in the transparency toolkit–are not getting answered.

The overarching assumption for our discussion is that there should be transparency at every step of the data lifecycle–from planning through publication. We should know…

- What are the data sharing agreements?
- What data collections are running and which have been axed?
- What are the standards for transparency around linked data?
- What are the standards for implementing AI/ML tools?
- What questions have been changed?
- What databases have been altered?
- What's being withheld from the public?
- What's being shared within the federal government (maybe inappropriately)?

…and more.

Guiding Questions

To seed the discussion, we asked six questions in each breakout group. The answers to those six questions are summarized below.

What are some examples of lack of transparency now?

Participants were able to identify numerous examples of lack of transparency today. Those examples, grouped thematically, include the following:

- Lack of clarity about what is being changed and now:
  - It's unclear what databases have been altered and what questions have been changed on surveys and forms because there has been no clear public notification. In many cases "last updated" dates have not even been updated
    - This includes revising data fields, codebooks, and even archival survey instruments on gender.
- Reduction in external communication:
  - Erosion of data user input, including the dissolution of many key advisory groups.
  - Census employees are not traveling, which cuts them off from data users.
  - CDC communication folks have been "RIFed" (RIF = reduction in force).
  - Paperwork reduction act staff (those who process the federal register notices and calls for comment) are largely gone in a lot of places.
- Reduction in reporting:
  - Some agencies are still producing microdata, but are no longer publishing tables (which reduces accessibility for some users).
  - Despite statutory requirements related to data collection renewals, not transparent when ceasing or no longer collecting data on a product.
- Reduction in oversight:
  - The Freedom of Information Act (FOIA) used to provide a way to force transparency. But now statutory response timeframes are not being followed. In some cases the entire FOIA team at agencies has been RIFed.
    - For example, we know the Pregnancy Risk Assessment Monitoring System (PRAMS) was shut down because state people leaked that info. At least one FOIA request has not been returned (a clear violation of public records law, but it's unclear what the appropriate next step should be).
- Knowledge about staffing and contract cuts and agency reorganizations has largely been through leaks, rather than through clear agency communication:
  - There is little to no transparency about which teams/entities still exist.
  - Some staff are being called back, which is adding to lack of transparency around what is happening (and makes it harder to encourage them to speak to the press, etc…)
  - Contracts changing and can't piece together why the amounts are less — are they cutting variables? Cutting somewhere else?

- Lack of clarity about many methodology details:
  - Unclear "how many ingredients go into the cake" to make the data (e.g. Population Estimates rely on Medicare and education data, but there are other less-well-known examples).
  - How are data being recoded to comply with Executive Orders?
- Lack of transparency on how data is being collected and used outside of its original purpose:
  - Examples include SNAP data for immigration enforcement, Justice Dept seeking voter roll data.
    - One issue on the sharing topic is direct grantees (e.g. Head Start), whose sophistication with data can be highly variable.
  - Is the Census Bureau being transparent when they release records?
  - What ways are our data being sold, given away, etc without our knowledge (to DOGE or others)?
  - Lack of transparency about data sharing agreements and information about who is sharing vs. who is resisting.
  - Even before 2025, how is it being collected, how is it being stored, etc.
  - When are certificates of confidentiality being breached? How would we know?
- There are existing gaps in understanding:
  - Local data users (e.g. local government staff, advocates, education administrators, etc…) may have low knowledge that the data they rely on are based on, or flow through, federal data infrastructure.
  - The understanding of what's collected federally vs. what is collected at the state. Discussions around SNAP data really bring this to light.
    - Non-governmental organizations matter as well — headstart, workforce programs, etc
    - Another part of the state vs. federal — how is your data being used now and what privacy protections are in place? "The government has your data" .. but which government?
  - Are the states talking to each other? (In education at least some folks are.)

One question that was raised is: "Who are the whistleblowers?" It is often unclear who is in a position to improve transparency and how to be supportive.

Folks are leaking to the press or to congressional staffers to create a public record that enables Congress and courts to take action. Unfortunately, most legal recourse requires demonstration of harm (which means, for all practical purposes, that action cannot be taken until after the fact). There are few (maybe no) preventive measures.

What was broken before 2025?

Participants acknowledged that transparency was imperfect, at best, before this year. One participant asked "was it ever that rosy?" Another noted that some of the challenge is institutional–it's hard to be transparent on complex data products/series. But there was universal

agreement that visibility into the workings of the federal statistical system is far murkier now than in recent decades.

Participants did identify several ways in which past efforts at transparency were inadequate or could benefit from improvement, including the following:

- Lack of clarity about what is being changed and now:
  - No additional issues noted.
- External communication:
  - Data producers think transparency is methodology. Is that the right framing? Others think of transparency as the ability to understand what government processes exist and why.
  - It would be helpful for orgs like APDU to create some standards for transparency.
- Reporting:
  - No additional issues noted.
- Oversight:
  - Federal Register Notice process was cumbersome and not always helpful.
- Staffing, contracts, and structure:
  - No additional issues noted.
- Lack of clarity about many methodology details:
  - Recoding of variables was (and remains) shrouded in some mystery. One participant noted that the Census Bureau used to recode same-sex couples into heterosexual couples.
  - Lack of clarity about imputation methods, such as imputation methods used in 2020 Census.
  - Need more transparency on data tradeoffs, such as reducing granularity to get greater certainty.
- Lack of transparency on how data is being collected and used outside of its original purpose:
  - Records management was (and remains)... messy.
  - In the past, people could opt out of having their data shared, but if our public data is increasingly private industry data, you often can't opt out. The only method of recourse is to change the law.
- There are existing gaps in understanding:
  - Themes noted above arose again here.

One bright spot noted during the conversation is that translation to the public was broken in the past, but might be improving now. Data experts are talking to media organizations, and the media is becoming more educated about reporting on this topic. Journalists are really thinking about getting as close as they can to the human stakes—what do these, sometimes wonky, data methodology and data governance issues mean for people, and what are the practical implications?

However, it is clear that there is still a long way to go. One participant asked "How do you explain in language that makes sense to the average person?" Others suggested more media training for data experts and data advocates.

What has worked well to provide transparency in the past?

Both in the "what has worked well" and the earlier question about "what was broken before" the Paperwork Reduction Act (PRA) came up frequently. PRA was described as a good framework, but one with some serious flaws. One participant referred to PRA as being "not a 10 out of 10 but maybe 5 or a 6." Some examples of positive aspects of the PRA (and associated Federal Register Notice/public comment process) include that there is a public comment process and that comments on proposed data rules are public. One participant presented the FRN on ACS disability questions as an example where the process worked well–advocates mobilized, the Census Bureau received more than 12,000 comments (most opposing the change), and ultimately the Census Bureau changed course based on public input.

However, most people (including agency staff) were not fans. Participants suggested that PRA really only works well if you're tracking a specific data collection. Scanning broadly is less effective. For that reason, several groups have started trying to fill the information gap–APDU, dataindex.us, and others are trying to make scanning more automated and then translate open calls for comment into suggested action.

The general consensus is that if we want to improve PRA, it needs an overhaul.

Other examples of what has worked well to improve transparency in the past include the following:

- Disclosure avoidance from 2020 census made code available so external experts could review, test, replicate, or research out new approaches and compare results. However, this was not particularly accessible to the "average" data user.
- Advisory committees helped with accountability
  - For example, Census Scientific Advisory Committee meetings were streamed online; Census Bureau responses were public.
- Opportunity Project – did a nice job with matching private sector and public sector; nice way to do work together.
- PRA process resulted in a log of what information agencies were collected and what they were asking. Complete but not user friendly.
- At one point someone developed a table (in Excel?) of what data had been used, what researchers were using it, and what questions were being answered. It was helpful, but no longer exists.
  - Need something like that for the "big picture" as well as nitty gritty details.
- To be helpful, a (new) tool would need to be user friendly, have search functions, and (ideally) be integrated in some way into data catalog work so you can see the end result of the data.

What would we want to see happen to improve transparency in the future?

Participants had several ideas for ways that transparency could be improved. Most, perhaps unsurprisingly, involved communications and outreach. Ideas ranged from making metadata more accessible, to improving internal documentation, to expanding external media and public relations efforts.

- Train people on how synthetic files are generated and how to use them.
- Improvements in metadata would go a long way toward alleviating some of the transparency concerns.
  - Metadata – fuel for AI to navigate the system.
  - You can get metadata now for many data collections, but it's not always understandable. Metadata may need a multi-tiered approach–codebase level detail, technical summary, and high-level/brief summary.
    - Example: Explaining "imputation" to your neighbor might confuse them, but a researcher needs the details on imputation. Communications strategies need both to be effective.
  - Also important to make documentation easy to find (not always the case today).
  - Metadata for public data is lacking, but administrative records may have no documentation at all. This is a problem that needs fixing.
  - Process and metadata transparency is huge and only the government can do that. External role is not to create the metadata but to push for standards.
- Conversations are needed on quality of data. What is quality?
  - Focus seems to be on singular number for an area/district.
  - For example, in development of county-level GDP, are the inputs considered? Focus is usually on granularity and not on the assumptions that go into producing the data.
- Improvements in internal standards for documentation would be helpful.
  - Example: IRS had an internal GitHub. But there are archaic systems–some things are still running on assembly language, others in COBOL.
  - Software has changed over time. It's important to know who is making changes and why.
- A lesson learned from the American Community Survey–telling people why they are being asked (each question) is important.
- There is an executive order to expand AI education across the K-12 system, there may be opportunities to help shape that curriculum in ways that improve data literacy more broadly
- Administrative records offer several opportunities for improved transparency:
  - Limitations on admin records aren't acknowledged, but the data are being incorporated into more systems as response rates fall.
  - For admin records to be meaningful, need a benchmark/comparison. (Triangulation is useful, but it's time consuming.)
  - Design surveys to provide baseline truth to hone the interpretation/use of admin records.

- ○ Need transparency around linkages.
- Increase external communications:
    - ○ Government staff never really talked to press, but because they hadn't been proactive in talking about the good stuff they were doing with data, it made it easier to tear apart. Getting out there in more positive and proactive ways would be good.
    - ○ Because agencies want to button down their communications they may never be the people to be talking to the media… So do they need to have the "talking" being done by outside people?
    - ○ Should the government agency being doing the reporting and doing the "spin" on the data? Or should outside groups be doing that?
    - ○ But who are the trusted voices? And what happens if that trustworthiness doesn't hold?

One unanswered question during the discussion is: To what extent do the fixes come from inside or outside of the government?

What tactics can we employ to make those things happen?

Across breakouts participants identified that having a "roadshow"--with general briefing information tailored to specific audiences–would be very helpful. For example, a briefing event for local governments (modeled after PRB's successful ACS on the Road series) could educate staff about federal data. By presenting in local communities, you can also learn about their needs. To support this work, APDU should develop a directory of conferences/industry groups to share what is happening with data (e.g., National League of Cities; NSCL; NACHO; National Association of Counties). Resources could include a slide deck that *anyone* could present across the country. Similar briefings for philanthropic organizations would be helpful.

During this discussion, participants shared that a recent webinar, hosted by MySidewalk, on 'Disappearing Data' attracted 350 attendees. In this case, the angle that data feeds AI attracted people's attention. This example illustrated that there is an appetite for more public education.

In addition, there are more ways in which the business community could engage to help improve transparency (specifically) and defend data integrity (more broadly). MySidewalk sets a good example. APDU should seek additional business ambassador(s) to help advocate for data for businesses and for workforce development.

Social media also came up in each breakout as a way to raise awareness and encourage advocacy. One group has a nascent "stat influencer" program in development, and that may be a useful place to start.

Humanizing the work is also key. From a strategic perspective, people want to be seen in the data, parties/influence, people want to be seen as on the cutting edge. Efforts may need to focus on examples that can resonate with everyone (e.g., NOAA and weather data, FAA and

flight delays, etc…) In addition, communications need to address people's fears. For example, once data are transparent, how will that affect me? Will it reveal undesirable things?

- FOIA - prevents release of trade secrets and pre-planning work
- In OK, now have longitudinal system - data stays with originator.
- Privacy - record linkages important. Other side is transparency, when Census started non-response imputation there was an uproar. Need development of acceptable practices.
- Think of data as the thing that provides insights not the thing that confirms bias.
- AI could be be employed to help with metadata/documentation.
- Leverage peer reviewers to employ some transparency standards.

What organizations are already leading (our role should amplify their efforts)

Participants listed numerous organizations who are doing helpful, related work on transparency who could be engaged as partners and/or whose work we should support. The list includes the following (in addition to APDU):

- American Libraries Association
- Council of the Inspectors General on Integrity and Efficiency (CIGIE)
- Federal CDO Council
- Committee on National Statistics (CNSTAT)
- Data Foundation
- My Sidewalk
- Project on Government Oversight (POGO)
- Sunlight Foundation - now defunct, but resources are still available.
- USA Facts.
- WIRED and any other media outlets who are doing a lot of work in keeping people up to date and tracking things.

One question that arose is: *Data advocacy has historically been a progressive topic. What conservative groups are data advocates?* Participants agreed APDU should explore how can we partner on topics that have obvious bipartisan appeal (e.g. privacy, data quality, transparency)?

National Archives or OMB should be leaders, but neither filling this role right now.

Where are there gaps that need to be filled?

There are many "big questions" that still need to be answered, including the need to define transparency and to develop transparency standards. Questions included the following:

- How do we define transparency?
- What do we do when transparency goes bad?
- What do we do when there is malicious compliance?

- When is transparency not enough?
- When is transparency not the goal?
- How should we be crafting transparency strategies?
- What should the standards be for transparency?

Participants expressed concern about existing mechanisms being insufficient to withstand the current environment. One participant described the need to wait for harm to be done before bringing a lawsuit as one key example and described the environment as being "sidelined" on immediate action. The ineffectiveness of FOIA requests (and what happens where there just is no response) arose repeatedly as one example.

Others noted that OMB is supposed to be providing oversight, but there seems to be little action. One possible role for advocates may be to call out OMB when regulations are not being followed.

Many participants noted that "leaks" to the press seem to be a primary path for transparency right now, but the potential "leakers" are often federal agency staff who could lose their jobs and may not have training or guidance on how to engage with the media, or even who trustworthy journalists might be. Several participants suggested "widening the pipeline" by encouraging recently-released staff to speak to journalists. APDU's role could include making introductions, hosting mixers, and/or providing media training to former fed staff.

Participants across all breakouts also noted a need for additional public education. Examples included the following:

- Need an information exchange to help amplify accurate, timely reporting.
- Increase general knowledge of CIPSEA (and what it covers vs what it doesn't).
- Increase public awareness of where (and how) they can refuse or revoke their consent to government usage of their data.
- Increase use of social media to help spread awareness of data privacy and transparency issues (e.g., "Ice bucket challenge" but for data privacy).

Participants also noted that education isn't enough. Groups need alignment on end goals, such as: Are there specific messages we want the public to hear? Is it sufficient that the public knows about data privacy (and breaches thereof)? Is it that they call their member of Congress?

There was a general sense from participants that public knowledge and media coverage can spur Congresspeople to take an issue seriously. A campaign to improve privacy and transparency–with stakeholders from across the political spectrum–may be helpful.

One participant recommended referring to the Frameworks Institute, rather than reinventing the wheel, for insight on how to craft a communications strategy.

Overall Action Items and Next Steps

Each breakout group identified specific action items, with themes that recur throughout, such as needing both short-term and long-term goals. The full list is presented below.

Short-term Actions

- Identify and collaborate with business ambassador(s) to help advocate for data for businesses and for workforce development.
- Develop a "roadshow" or "plug-and-play" presentation that anyone could give that can be tailored to
  - Local governments
  - Philanthropic organizations
  - Industry groups / associations

Long-term Actions

- Stat Influencer idea for Tik Tok to influence younger generation
- Catalogue and start to work through the hard questions:
  - How do we define transparency?
  - What do we do when transparency goes bad?
  - What do we do when there is malicious compliance?
  - When is transparency not enough?
  - When is transparency not the goal?
  - How should we be crafting transparency strategies?
  - What should the standards be for transparency?

2025 Data Integrity Summit - Day 1 Summary

The federal statistical system has been struggling with budget constraints for several years. In the first six months of 2025 threats to the integrity of the system increased substantially, including cuts to staffing and contracts, data removals, revisions to data collections, and delays and missed deadlines in the publication of key information. Day 1 of the Data Integrity Summit brought together data experts and users across the U.S. to strategize about recovering from the current disruptions and rebuilding an efficient federal data infrastructure for the years ahead.

Recurring Themes

Although breakout discussions were structured around several different hypothetical (but possible) data changes, there were several common themes across the breakouts.

Transparency and public trust are paramount

Across all breakout sessions, transparency and public trust were identified as critical considerations for any data system changes. Clear communication (from planning through publication), broad public access to both data and documentation, and robust protections against misuse are essential to maintain public confidence.

Data quality is key

A recurring theme was the challenge of tracking and ensuring data quality in new or consolidated systems. This involves defining new metrics, conducting robust testing (e.g., parallel runs for surveys, comparing administrative data to traditional methods), providing error metrics, and publishing clear documentation on data methods and limitations.

Education is needed

While participants were all deeply knowledgeable about threats to the federal statistical system, many data users are a step (or more) removed from knowledge about data infrastructure and the current threats to data access and quality. More education is needed at several levels including among elected officials, membership organizations, and data users.

Privacy protection is critical

Given the increasing use of linked survey and administrative data, as well as changes in the way the government is using data, many participants called for enhanced privacy laws, strict data use agreements, and robust oversight mechanisms. This includes ethical/privacy review boards, community-informed oversight, and increased penalties for breaches to protect against misuse by government agencies or third parties.

Overall Action Items and Next Steps

There are specific actions identified for each discussion above, with themes that recur throughout. APDU's key action items fall into categories that mirror the key themes above.

Transparency and public trust

- Push for transparency at every step of the data lifecycle–from planning through publication.
- Consider using the advisory committee model as a mechanism for keeping up pressure for transparency in decision-making.
- Propose standards for transparency around linked data, such as (but not limited to) minimum reporting standards.
- Propose standards for transparency about the implementation of AI/ML tools. Most of these are "black box," so there needs to be robust processes (like making training data available) to ensure transparency.
- Under this umbrella, there is also a need to support additional research into the implications of ongoing changes such as the use of AI and ML, use of administrative data, consolidation across surveys, implementation of privacy-enhancing technologies, and the associated and implications of changes.

Data quality

- Ensure that any changes to data collection are based on a scientifically rigorous implementation strategy (and technical plan to permit benchmarking, quality assessment).
- For AI/ML learn more about what has already been done within the federal statistical system and what are the lessons learned that could inform future ethical use.
- Also identify the goal(s) of any changes–are the goals to reduce respondent burden? Reduce cost? Improve accuracy? Improve timeliness or granularity? Improve privacy? Be prepared to demand that those goals are clear at the outset and that there is evidence that the change(s) will support the desired outcome(s) *before* implementation.

Education

Develop a strategic communications plan, that may include:

- Segmentation by audience
- Risk assessment (will communications help or hinder progress)
- Bilateral meetings with left, right and center think tanks to find common ground.
- Research into what messaging strategies are effective in this environment, and if we find nothing, hold focus groups to learn more. Subsequently, conduct training sessions for the media and data stakeholders on best practices.

Specific actions may include:

- Hold a convening and invite agency staff so we can learn what they need and want. Also to connect federal agency staff with data users so we can turn data users into data advocates.
- Compile primary uses for key statistics/data elements so that we can reference these as we develop messaging; flesh out use cases for specific audiences. (*Editor's note: Work of this type is underway–APDU's role may be to support and uplift those efforts.*)
- Take APDU out to the streets, not just a yearly DC conference.
- Continue to host events where data users can learn about the latest state of affairs, share information, and strategize.
- Conduct outreach to partner and peer organizations (like NASBO, NCSL, and others) to raise their awareness of ongoing changes in the federal statistical system.

Privacy protection

- Outline a plan for how enhanced privacy standards and oversight would function (work with Representative Trahan's office).
- Support efforts from community-led organizations to make sure all voices are heard.

Breakout 1: What if most (or all) surveys are collapsed into an "omnibus survey?"

What are the risks and benefits of survey consolidation? What are the implications for data users?

Consolidating many surveys into one (which would potentially involve an incredibly lengthy questionnaire) presents significant challenges. A longer survey, like an omnibus survey, is likely to lead to greater respondent attrition, especially in small populations, and could undermine trust among participants if sensitive topics were included. A consolidated approach also makes the survey a vulnerable target for critics, potentially leading to its politicization. In addition, changes to survey administration would lead to substantial breaks in critical data series, making trend comparisons difficult or impossible. It's also impractical to combine all current surveys into one due to diverse subject matter and specialized expertise required for individual surveys, leading to lost content and coverage. A middle-ground position of fielding a small number of "minibus" surveys (e.g., centered on health, education, economy, etc.) may have promise.

The potential merits of massive survey consolidation include basic existence: if the alternative is that survey content would be lost entirely, consolidating questions into a larger "omnibus" survey may be better than having no data at all. There could also be benefits from name recognition (today, ACS is better known than NCVS, for example). In addition, having data linked across multiple topics may lead to insights that individual surveys have not been able to uncover in the past. Thinking in terms of an omnibus survey may also help rethink the entire data collection enterprise, identify solutions to data linkage barriers, and improve efficiency in data collection. There may be efficiencies gained from having a common core of demographic questions, for example, rather than fine tuning question wording across multiple survey instruments.

Regardless of the mechanism, any survey infrastructure needs to be able to support data for subpopulation analysis, especially for small subpopulations. Any consolidation across surveys would require clear criteria for the included topics along with an assessment of which topics might lead to respondent dropoff and how to mitigate those challenges. The sampling frame would also need to be carefully considered. A survey focused on education needs a different frame (children) than a survey focused on workforce (adults), and a survey of individual characteristics (such as health) needs a different frame than one focused on family or household characteristics (such as household composition and income).

The long-standing problem of whether people are filling out a survey accurately would remain, and continue to be difficult (or impossible) to probe.

How can we mitigate the risks we just discussed? How can we ensure potential benefits are realized?

Trust is essential for a consolidated survey to succeed, so maximum transparency would be essential–that includes transparency in access to the data. The more that data collection could be portrayed and protected as anonymous/private, the more trust could be built.

An omnibus survey will need clear legal authority and stable funding authority to get started–it could not rely on the current patchwork of authorizations.

Would we need to revisit the idea of substantially paying survey respondents and/or providing a strong incentive to participate in a more massive survey collection? (Tying survey completion to a payment or tax incentive, however, might undermine other work on privacy.)

An omnibus survey cannot cover all topics of interest nor all questions. Would a series of "minibus" surveys be more appropriate?

Any consolidation effort would also need to identify how administrative records would support/supplement a survey. Need to keep in mind inherent challenges of measuring an inherently mobile population. For example, vital records/admin records may be strong at birth/death landmarks, but less good at other transitions. An omnibus might be poor at picking up within-population shifts/transitions. To be effective, omnibus resources may need to be steered toward small/sensitive/undercounted populations.

How would we track data quality for data users?

A key problem for tracking data quality if an omnibus survey were implemented is that consolidation would necessarily involve eliminating measures previously used as a benchmark for comparison. Ideally, any new survey structure would be run in parallel with existing structure for a period of time to allow comparison.

Tracking user experiences would require user group engagement, and that's difficult to do in an omnibus, please-everyone climate.

A consolidated survey may result in different responses than smaller, more tailored surveys do. For example, the American Indian / Alaska Native population has the most fluid claimed membership in surveys, and response differs by the specific context of the survey; those kinds of biases might be harder to peg in an omnibus/minibus environment.

An omnibus survey may also change respondent behavior in unexpected ways–such as a sensitive question on one topic leading to different responses on another. Considerable research and testing would be needed to identify those interactions.

What are the lessons from this exercise that should inform the federal statistical system of the future?

- There will be information loss/coverage loss moving to an omnibus environment, inevitably; at what point is getting something (high-level indicator for main groups) better than getting nothing (cutting programs)?
- Omnibus strategy (core sample, topic modules, control over rethought/prioritized content) might be a useful future direction for rebuilding the enterprise in 10–25 years time, but not really for short term
- Who needs to be in the room to decide what's needed, sort through legal needs for information, articulate content and governance for the content of the survey?

Action Items and Next Steps

- Steer conversation toward a set of minibuses (health, education, demography, economy, justice) that might be workable, as well as scientifically rigorous implementation strategy (and technical plan to permit benchmarking, quality assessment).
- Think about the extent to which the survey would have to be farmed out to states (knowledge of uneven patchiness in coverage could be better than nothing, but could raise new implementation concerns).
- Continually have these kinds of conversations (landscape changing so regularly).

Breakout 2: What if the 2030 Census becomes an "administrative records census?"

What are the risks and benefits? What are the implications for data users?

One of the biggest risks is the timeliness and completeness of administrative records. State administrative records vary in quality, and administrative record providers may not have sufficient staff to support providing data either using traditional mechanisms or with new privacy preserving technologies. Federal records tend to be biased towards higher income people and households (such as people who file taxes, receive Social Security benefits). Most administrative data sources miss people experiencing homelessness.

Even the most basic administrative data that underpin the current "door-knocking" census framework–the master address file–misses some addresses. Administrative records may have better coverage of these households Within addresses, some households have multiple families or unrelated individuals. Administrative records may also rely on a mailing address (like a PO Box) that does not align with Census residency rules.

A shortcoming of administrative data sources is the absence of key characteristics. For example, race and ethnicity data are not included in tax records. Some benefits programs have race/ethnicity lack detailed information. Also, such information may have been recorded by an administrator rather than self-reported or may be missing. Most administrative data sources have not yet been updated to align with SPD-15 reporting standards.

In addition, right now, record deduplication methods are "black box." There would need to be robust, transparent testing to demonstrate that the processes are working.

There are concerns that a move toward using administrative records for census purposes could drive more people "off the grid," making people avoid public benefits and public school. This may be especially true in immigrant households or households with transgender members.

There are concerns that administrative records might be dismissed as an impermissible method of counting the population from a legal standpoint.

Ultimately the biggest risk is that changing methods may undermine trust in the data and lead to a "failed" census. At a minimum, all stakeholders would need to understand the downstream effects on data, since the decennial census is used for the benchmark for many characteristics besides just population (age, household structure, family, race/ethnicity, etc…).

There are, however, also potential benefits including improved speed, reduced cost, and–at least for some populations (like young children) potential mitigation of existing undercounts. For example, apportionment only requires state totals, which administrative records could probably produce fairly well.

If an administrative records census were effective, it could potentially be done more than once a decade and could lead to improved population estimates between censuses. An administrative records census could also potentially expand the characteristics included in the decennial count without increasing respondent burden.

Increased research and transparency about the potential uses of administrative data could illustrate that the data and challenges are complex–and could help mobilize resources to advance the field.

How can we mitigate the risks we just discussed? How can we ensure potential benefits are realized?

Any system changes would need to move very quickly. Participants suggested that transparency must start "NOW on everything." Transparency is not just about what the bureau is doing but also about their reasoning.

One potential check on data quality would be to implement a method through which people can check their own data. It would be logistically challenging, but could help both with trust and with accuracy.

Improving state records would be beneficial before using them for the census. This may be unlikely in the current state of government, but high-quality state data would be critical to an accurate count.

Enhance privacy laws to make it illegal to use the data for specific uses.

Eliminating block data was also raised as a potential mitigation against data quality and privacy concerns.

How would we track data quality for data users?

First, we would need to define data quality metrics - validation rules, establish standards (eg characteristics). This could include measures like imputation rates for detailed race/ethnicity Stakeholders could look to other datasets (non-federal data) to help design a data quality framework.

Accurate coverage measurement would be critical. One suggestion was to conduct the Post Enumeration Survey (PES) with a sample of people from the admin data, and use traditional survey methods to see if self-report from those people aligns with the administrative records. But others noted that we need to move away from our old ideas that enumeration and PES are sufficiently accurate. If we question the accuracy of new methods, that should be pegged against whether it is "more or less" of a risk to data quality compared to current issues. For example, PES does not get at missing people found in an admin rec census nor does it capture quality for young children accurately.

Tests could start now and should include key insights such as how well do AR reflect the population and where you need more/diff data. This flows into a need to model various sources/combinations of AR data, especially for small areas. Need to understand what data to use and when.

Need more transparency - even about current estimates.

Part of the problem is that we don't know who all users are now or how data are being used. How do we convey quality/accuracy to people using just one number, especially for people who are not looking at MOE now?

Need to understand the downstream effect on data products. Many things rely on decennial census - eg state projections, area profiles. Quality changes to decennial have broad effects across the data ecosystem.

What are the lessons from this exercise that should inform the federal statistical system of the future?

- Agencies need to take more of an interest in their agency's administrative records and make these inputs usable for other agencies.
- Need to define terms like accuracy. Topline numbers are probably going to be pretty good. In fact, AR is likely better than traditional enumeration. But for characteristics it will not be. The unit of analysis matters a lot.
- Why use AR - just because it's cheaper? Is it actually cheaper? Even if it is, should that be the deciding factor?
- How will this affect how agencies create samples for other surveys? how will AR be able to compensate if a traditional decennial benchmark goes away (impact on downstream products)?

- How data will be shared?
- 2030 is going to be a both, not an either/or (trad and AR); remember that AR are collected for administrative purposes (ie building permits have reporting issues)
- To do this well, we need a united states - need everyone working together, trust for data sharing
- Need Privacy-Enhancing Technologies (PETs) (federated data, private set intersection)
- Expect declines in public trust
- Enhance privacy laws, transparency. For example, laws that put penalties on anyone, not just statistical agency staff, for breaching confidentiality. Such legislation is needed to protect against ICE, corporations, etc. The concern with reidentification from public use files is second-order compared with the concern of government agencies invading stat agencies files.
- When we use census data- its persons and household level, etc- are we going to be able to attribute household level characteristics to persons and vice versa using data derived from an AR only census?
- Are there international examples we could look at to help address some of the questions we have and/or use as templates?
- Detailed information about people can be reported differently by the same person in different contexts, so the administrative data is going to reflect the relationships that people have with those administrative units and their trust and use of those data.

Action Items and Next Steps

- Push for transparency:
  - Get Census to announce their plans for use of administrative records and/or meet with Census staff to advocate for them to make a decision as soon as possible.
  - Also get states to push for transparency. The Census Bureau's Redistricting Office is starting to meet with states in person this summer. APDU could help educate and organize state redistricting liaisons before those meetings as a way to push for transparency. State legislatures and governor's offices will be involved in those meetings.
  - Consider using an alternative advisory committee model (as though they were still formal bodies).
    - Census Quality Reinforcement Task Force will be convening Census Scientific Advisory Committee members in Sept. Setting agenda soon
    - Union of concerned scientists produced a toolkit on how to do an outside Federal Advisory Committee Act committee.
  - Have the Census Bureau define in-office enumeration - they are likely to say "we will not do a full AR census" but need to prompt them to say more
  - Offer an opportunity for Census Bureau to use APDU events for input (staff want to produce high quality, useful, and accurate data, but communications have been stymied)

- APDU sending letters to diff parties who can push CB to be more transparent (to CB, oversight, congress, flood the field)
        - Use FOIA, if necessary
    - More research is needed into…
        - Technical options (federation, PETs)
        - Cost (is it actually cheaper?)
        - Find ways to enlist community help with research that needs done.
        - Need quality assessment of the "Best Race" file at Census, who does it cover, how does it handle discrepant responses (across sources and over time), does it include detailed R/E?
        - Are there international examples to inform US options and decisions?
        - Implications of admin data on citizenship in census
    - More education is needed
        - Need a short APDU white paper describing all the ways various private sector/ML/state data are benchmarked to the census
        - The advocacy community as a whole may need to know more.
        - Work with NASBO (budget people, National Association of State Budget Officers), NCSL–reach out to groups, being them up to speed (bit by bit)
        - Prioritize outreach to state and federal geospatial groups
        - Raise all these issues with non-data people (who are not following discussion but will be impacted if products deviate from past, won't understand limitations)
    - Congressperson Trahan issued an RFI about updating the Privacy Act: have APDU get a briefing on that from staff
    - Need a strategic communication plan, stakeholder engagement plan
        - Lead up and engagement differs by group
        - Be aware of the risk of getting people together - damage instead of progress w/ the CB
    - Take APDU out to the streets, not just a yearly DC-based conference

Breakout 3: What if the government tries to create a disease registry from administrative records?

What are the risks and benefits? What are the implications for data users?

Key risks from using administrative records to create a disease registry are threats to privacy and use for non-statistical/non-health purposes. Linked data could reinforce harmful stereotypes or be used in a punitive way, as with HIV status. Without adequate guardrails, a registry could be used in ways that harm the population such as through denying access to healthcare or employment. Health status could be linked to receipt or denial of benefits, such as refusing safety net benefits to someone with a substance use disorder. In the past, health information was used in nefarious ways including forced sterilization and imprisonment. All of these risks, and more, arise if administrative records are used to expand population surveillance.

In addition to accurate linkages causing privacy concerns for people with a given health condition, false-positive linkages introduce risks for *all* people. In addition, linked data may miss some people who might benefit from any advantages derived from the data. Automatic methods (e.g. machine learning) might be required to identify who is included in a registry, which will include errors (especially errors that are racially biased).

As evidenced by the reaction to a recent "autism registry" announcement, the use of data for purposes other than what it was collected for greatly reduces public trust–even if intentions were good. Suspicion then may lead to reduced utilization of healthcare, which may additionally shift access from primary care to emergency care–which increases cost and burden across the healthcare system.

Despite clear and worrisome risks, there are potential benefits for health research and evaluation. Linked data could improve governmental ability to identify people who qualify for a supportive program who aren't enrolled. Registry data would improve potential for:
- Improved cross-agency coordination.
- Better monitoring the burden of disease.
- More precise targeting of treatment and prevention.
- Easier identification of causes (e.g. environmental exposures) and address the cause with policy (or hold accountable through litigation, if appropriate).
- More efficient and appropriate allocation funds and other resources.

Data linkages could also save money and reduce respondent burden by not needing to gather primary data. Unfortunately, in practice this has not yet panned out. For example, there have been questions about whether a vision module is needed in NHANES study, given ophthalmic data, but so far the survey data are still needed.

How can we mitigate the risks we just discussed? How can we ensure potential benefits are realized?

Since most of the risks related to privacy leakages and misuse, mitigation focused on those issues and included the following:

- To mitigate and/or address errors, allow individuals to request removal from database and/or correction of record? (A challenge here is whether or not people would even know if they were "on the list.")
- Implement strong data use agreements—clear purposes, retention limits, access protocols.
- Implement informed consent procedures share related to personally identifiable information.
- Develop a risk mitigation framework.
  - If the risk exceeds a given threshold, store data in aggregate to prevent incorporation into registries.

- ○ Regardless of the threshold, use privacy enhancing technologies to prevent identifiable linkage.
- ○ Consider limiting linkage and monitoring to common conditions, like heart disease and diabetes. (Threshold could be linked to a percentage of medical spending?) More common conditions may result in lower risk of data misuse.
- ○ Ensure that the communities affected are involved in setting data policies.
- ○ Increase penalties for disclosure and make sure penalties are enforced. (Is it time to have data and linkage live outside of the executive branch? Or at least with stronger oversight from a different branch?)
- Educate the public and legislators about the risks and potential harms of registry data.
- Clearly identify who's in / who is missed–compare linked data to other sources (such as surveys) to help identify any under (or over)count problems.

Review and oversight are also critically important. The group came up with several strategies including:

- Implement ethical/privacy review boards to evaluate and monitor use.
- When linkage is proposed, have oversight/review that is informed by the community affected–don't just rely on the stated reason for the request. What looks "helpful" from one perspective, may be harmful from another.

How would we track data quality for data users? How would we monitor privacy?
Data would need robust testing to ensure linkages are accurate and also ensure those who work with the data understand that no linkage is ever perfect–there will always be false positives/negatives. There would need to be clear documentation and understanding of limitations to how data can/should be used.

The responsible agency would need to be transparent and publish clear information about how the registry was created, how it's used, what privacy techniques are applied, and how a person can opt out if they choose. CBOs and "trusted messengers" could be helpful outreach partners on dissemination and trusted messengers. Any documentation should use plain language that is accessible to the public.

The responsible agency should also routinely publish data quality reports that summarize important metrics like linkage errors, etc.

What are the lessons from this exercise that should inform the federal statistical system of the future?
Need enhanced privacy standards and oversight

- Need ethical/privacy review boards to evaluate and monitor use
- When linkage is proposed, have oversight/review that is informed by the community affected. (The NCVS letter about gender is an excellent example of community-led advocacy.)

- The good actors / bad actors problem is challenging: How do we build in mechanisms so that the system is resilient in the wrong hands but is still beneficial?

Need clear communication about data quality and appropriate uses:

- Publish routine data quality reports that summarize metrics like linkage errors, etc.
- Clearly identify who's in / who is missed–compare linked data to other sources (such as surveys) to help identify any under (or over)count problems.

Develop a risk assessment framework for when we advocate for data access and/or linkage and when we resist it because the risk is too high. For example, providers need to report accurately, but also need to know that they aren't putting their patients at risk. The risk framework needs to be flexible enough to cover different data collections–healthcare records, surveys, and admin records all have different risk profiles and considerations.

Action Items and Next Steps

- Outline a plan for how enhanced privacy standards and oversight would function
- Propose standards for transparency around linked data (minimum reporting standards?)
- Support efforts from community-led organizations (like the NCVS letter)

Breakout 4: What if ML and/or AI became the default for inferring data when it's not collected?

What are the risks and benefits? What are the implications for data users?
The group identified lack of transparency as a substantial risk–AI and ML tend to be black boxes. In fact, transparency is antithetical to many AI tools. Some companies have been known to litigate in order to protect what their tools are doing. This means:

- Data users don't know what is happening in the modelling
- Data biases may be exacerbated
- Data may be less representative
- Reproducibility may suffer

AI and ML lack a human oversight element–someone looking inside the black box. Processes need to be well-documented to be beneficial to the public.

There are also potential benefits to broader implementation of AI and ML tools in public data. They may be faster and less expensive. They may reduce respondent burden. They may allow production of data that are more granular and improve the ability to tease out trends that might otherwise remain hidden. And they may subvert some *human* biases.

It will be important to know which tools the federal government ends up using, the origins of the data, personal consent to use the data, and ways to mitigate any bias.

How can we mitigate the risks we just discussed? How can we ensure potential benefits are realized?

To mitigate risks:

- We need guardrails around how/who/what is being used, such as:
    - Identifying the necessary documentation and resources to ensure processes are transparent and reproducible. For example, make the training data public. That way, anyone can run models and try to reproduce results.
    - Implement protections around the use/non-use of certain data sources.
    - Respect existing data ethics principles - cannot abandon ethics in favor of new technologies.
    - Enact governance structures for the Federal Statistical System that support ethical use.
- Inform legislators about risks and benefits.
- Make sure students, young people, are educated on AI/ML implications (specifically people outside of the data community). Public awareness is key.

Any policies or practices that are in place should be geared toward ethical use and building/maintaining public trust.

How would we track data quality for data users?
- Need transparency around data quality, which may require developing and reporting out new measures
    - Need to know margins of error for data sets
    - Make sure measures are understandable to regular people
    - How are we measuring missingness? Coverage?
    - What was done, who did it, how was it done, and what are those measures that we need for validation?
    - Use benchmark datasets to look at and evaluate various measures of bias (non-response, selection, linkage).
- Run simulations: under other assumptions how do quality measures change?
- Re-validation/calibration for updating data quality measures
- Acknowledge trade-offs: infusing noise (lowering quality), increasing uncertainty

What are the lessons from this exercise that should inform the federal statistical system of the future?
The key theme in the discussion is that the system is not going to look the same moving into the future. It will look different, operate differently, but hopefully principles and practices will be the same or very similar.

- We need processes tailored to AI and ML to ensure effective guardrails.

Staffing and knowledge are key considerations. We need knowledgeable people in the system to appropriately apply methodologies. This may include increasing collaboration with outsiders

(both on technical and consumer sides) to bring new ideas and new technologies into existing work. We also need to cross-pollinate existing knowledge on how to model statistically (how to evaluate, create standard errors, evaluate linkage keys, etc.). Existing knowledge should help inform future directions. At the same time, there has been a huge loss of institutional knowledge in the last few months. We must bring in new people with no knowledge and invest in educating them.

There is also a need for education broadly. Discussion participants shared examples of situations in which smart data scientists demonstrated a severe lack of knowledge around the federal statistical system. This goes to illustrate the need for further education to push back on losing institutional knowledge.

Some survey data will likely still be needed. Not all questions can be answered from administrative datasets and records, and there is selection bias in administrative data sources. This means we need to think creatively and strategically about what data elements are important to collect from the entities that we need to collect from.

Last, but not least, there are hype cycles around AI and ML–with advancements in technology over promising and under delivering. Any move forward needs to be clear-eyed about what is possible (and what is not).

Action Items and Next Steps

- Do we know what, within the Federal Statistical System, has already been done? What moves have already been made along these lines? Where can we start with allies/champions for transformational culture changes? How do we quell fears?
- We need to decide on a structure. What does that look like? Are we in agreement on what structure to use? Are we including voices from various places?
- Must identify existing staff with the needed skill sets. We must identify mentors for "onboarding" new staff.
- Begin with the 30,000 ft level.
- Who are the decision makers in these next steps? What responsibilities fall into which lanes?
- Concrete steps might include convenings, putting together list of recommendations, written products - Inform and influence
- What are the concrete goals of AI/ML solutions? What are we solving for? How do we get these tools in front of people to vet the systems before wider distribution

Breakout 5: How do we better tell the story of why federal data are important?

Relevant data

Participants listed a number of relevant datasets and statistics they use in their work, and we can group the responses into the following categories:

- Social sciences
    - Criminal justice
    - Census and ACS
    - Economic data and statistics
    - Child welfare
- Education
    - IPEDS
- Health
    - Vital statistics

Given the composition of APDU membership and Summit attendees, it is not surprising that social science and health data products were particularly important.

During the discussion, participants emphasized the importance of having data for small geographic units since federal data is one of the only ways to obtain reliable data for such entities (even if uncertainty estimates are higher for small units).

One participant mentioned Ron Jarmin's talk at the ACS Data Users Conference, where Ron argued that we should stop thinking about data as surveys and estimates. Instead, we should think about what we want to get from the surveys and estimates (e..g, we need better measures of employment or income). This is a provocative statement that may require more discussion / argument.

Value of federal data

Participants highly valued federal data, and their reasons / explanations can be grouped into the following categories:

- Federal data are freely available to everyone in the world.
- They provide the stable, reliable foundation upon which other data products or research is based.
- Federal statistical agencies are incredibly transparent about how they collect, prepare, and publish data (sometimes to a fault).
- Federal data have the scale and scope to support the measurement of small geographic units and/or small population subgroups

In our discussion, the interconnectedness of federal data products came up. One participant mentioned that GDP estimates take in 200 inputs, with 70 coming from statistical agencies other than BEA. I mentioned the importance of population estimates in other data products, including the CPS, ACS, and the National Health Interview Survey. Even people with federal statistical agencies do not have a sense of how interrelated these data products are.

Compelling and not so compelling types of communication

Participant input and the discussion was wide-ranging, with a focus on communication that works well or is compelling. We only had a few examples of what does not work well (locking up methodology discussions in dense PDFs, ignoring criticism of data, even using the word "data", calls to save or restore dataset X have fallen flat).

One participant made a good argument that simple one-pagers with highlighted take-home points are effective. They pointed to Andrew Reamer's Counting for Dollars project as particularly compelling, especially for state and local officials. Explainers from USAFacts fall into this category as well.

Communication that makes the business case for federal statistics is also compelling. Tying federal statistics to economic development projects can help convince people of these data's importance, and pointing out what may be lost if federal data were eliminated or made less trustworthy.

During the discussion, we pivoted a bit towards who is the audience for different types of communication. One participant mentioned foregrounding how the data helps advance the goals of organizations as opposed to foregrounding the data themselves. We had another participant emphasize the idea that money is compelling - "how much money will we lose if we don't do something"? That was a common topic in this person's experience advocating for federal data.

Messaging, messengers, and audiences

With respect to messaging, we discussed developing specific calls to action since that would help us identify our target audience and the best messenger for that audience. If the call to action to fund federal statistical agency X, our target is Congress and a good messenger would be a constituent or local person who would be harmed by cuts to agency X. Additionally, there is a need for multi-modal messaging to reach a larger audience.

With respect to audiences, participants identified the following who may benefit from outreach:

- Congress and the administration (and the people they will listen to)
- Senators representing rural areas
- State and local officials
- Business owners
- Parents

We need to identify the people who have the hard/soft power or decision-making authority to enact the change we seek.

With respect to messengers, participants identified the following:

- Local business people who may be hindered by a lack of reliable data (harm to economic development)
- Rural constituents who may be harmed by lack of data for their smaller places, especially related to economic development and health care infrastructure
- State and local officials who use federal data to advocate for economic development projects in their localities

Lessons learned from our exercise

The group agreed that we need to develop priorities and then stay as focused as possible on those priorities. We can't try to boil the ocean, so we need to focus on a handful of items. We also need to tailor our messaging for specific audiences. We cannot create a "one-size-fits-all" approach because it will not resonate with audiences. Our messaging needs to meet people where they are.

News hooks matter, and we should not let a good crisis go to waste. If something happens, we should be prepared to deploy messaging either cheering or decrying that thing.

One participant mentioned a particularly effective statement related to higher education. That statement was so effective and powerful because it was issued jointly by two groups that were at odds with one another. Thus, coalition building among organizations that may not agree on much could yield the best results.

Action Items and Next steps

Participants listed a number of steps we could take to continue developing messaging, identifying the best messengers, and targeting specific audiences. Those steps include:

- Bilateral meetings with left, right and center think tanks to find common ground.
- Hold convening and invite agency staff so we can learn what they need and want.
- Compile primary uses for key statistics/data elements so that we can reference these as we develop messaging; flesh out use cases for specific audiences.
- Connect federal agency staff with data users so we can turn data users into data advocates.
- Collect information about what messaging strategies are effective in this environment, and if we find nothing, hold focus groups to learn more. Subsequently, conduct training sessions for the media and data stakeholders on best practices.