Poppy MacDonald Testimony
USAFacts

Good afternoon, Chairman Kilmer, Vice Chairman Timmons, and Members of the Select Committee on the Modernization of Congress. My name is Poppy MacDonald, and I am the president of USAFacts. Thank you for the opportunity to speak today. Our organization appreciates the committee's dedication to open data and commitment to leverage it as a tool to strengthen the lawmaking process and improve policy. I'm grateful for the opportunity to be here alongside Dr. Nick Hart; as members of the Data Foundation, USAFacts shares the organization's commitment to improving government and society by using data to inform public policymaking.

USAFacts is a nonpartisan, not-for-profit organization dedicated to providing the American public comprehensive and understandable government data. Our organization empowers American citizens, and you and your fellow elected leaders, to make data-driven decisions about the issues facing the country. USAFacts does this by providing unbiased facts by standardizing data from federal, state and local government sources, and presenting it in a clear and simple manner. The data we use is consolidated from 78 federal agencies along with sources from every state and region nationwide. We don't attempt to influence public opinion in any way; we consolidate government data to support serious, reasoned, and informed debate. We exist to help all decision-makers, as well as the public, access reliable information and empower them to make policy from a strong foundation of trusted facts.

Now more than ever, easily accessible data is essential to ensure that not only our elected leaders, but all Americans can make informed decisions. In the face of increasingly complex challenges in our country, USAFacts believes in a thriving democracy supported by undisputed, trusted, comprehensive government data. USAFacts was created out of a clear societal need for accurate and accessible data from federal, state and local governments. **We believe the facts deserve to be heard.**

In 2015, our founder and former CEO of Microsoft, Steve Ballmer, wanted to use his personal charitable giving efforts to help lift children out of poverty. He tasked a team of financial analysts with researching impartial data around the government's own efforts to determine where current programs are targeted and where there are additional needs. The analysts thought the research would take a matter of weeks to compile, but it actually took six months. The information these experts were able to compile from the government was disparate, difficult to access, and outdated.

Steve asked an important question: Why isn't the same standard for metrics and data that businesses are required to report to shareholders applied to the impact and progress of government programs and initiatives? Aren't we all shareholders in our country who deserve the same access to transparent, publicly reported data? This knowledge gap forces citizens and policymakers at all levels into a difficult position, where they are called upon to make major

decisions with incomplete or dated information. In light of this challenge, USAFacts was created as a free resource for all Americans to have access to trusted facts.

We maintain a regularly updated website, publish an annual report on combined federal and state spending, revenues, demographics, and outcomes, publish an annual 10-K for the government modeled after the SEC Form 10-K that public companies must file, and share contextual content in social media channels.

Without the government actively sharing the vast trove of data it collects, Americans are left without an authoritative and unbiased source to inform their conversations, their opinions on policy, or their votes.

There are over 90,000 state and local government bodies in the United States. Few have standardized means of reporting information to agencies, Congress, and citizens. Agency data is often siloed in the process of collection, analysis, and presentation, resulting in confusion and duplication of data collection. Administrative and statistical data is often collected and analyzed for the sole purpose of implementing programs for one agency rather than use across agencies.

At the start of the COVID-19 pandemic, federal and state leaders lacked standardized data to inform crucial decision making. USAFacts combed through state and county reports, all of which used different reporting methods, to create a standardized view of daily virus cases and deaths. Not only did the American people need it, we have become a go to source for many levels of federal, state and local government, as well as publicly traded companies and not-for-profits, to provide comprehensible real-time COVID data. We filled a key need in the public awareness effort, but COVID-19 will not be the last challenge this country faces.

Successfully modernizing Congress to prepare for our current and future challenges depends on members having unencumbered access to reliable data on government programs and outcomes. One of the principal reasons for USAFacts' founding was to allow Americans to use data to make decisions about the future of the country in the same way that executives use data to make decisions for their companies. Modern businesses rely on robust, timely data to make strategic decisions, and a modern Congress should be empowered to do the same. However, Congress faces the same challenges in accessing and using government data that USAFacts and the American people face. Congress should work to reduce barriers to accessing government data, both as part of its efforts to modernize and to further avail data to the American public. Again, we appreciate all of your efforts on the committee to advance this need.

To support the vision of a modernized Congress that makes decisions based on data, USAFacts advocates for more open, timely, and detailed data from our government. We continue to promote recommendations that individual agencies can implement to mitigate access issues, including:

- **Ensure data is timely, complete, and accurate:** Some government datasets are several years old, making it difficult to see the impact of spending and outcomes of programs
- **Create easy ways to access and understand data:** Offer data accessibility standards that make all datasets easy to access and analyze by all individuals, regardless of technical expertise

- **Ensure datasets are contextual and relevant:** A single number or percentage point can be misunderstood without context. Provide numbers with historical context whenever possible
- **More cross-agency and cross-government collaboration and standards:** As we learned from the COVID-19 pandemic, states report data inconsistently and with differing definitions, standards, timeframes, and visualization platforms. Mandate and fund data collections with standardized formats to improve data pipelines
- **Make the data and collection processes more transparent:** Understanding data availability and format by agency would be useful to support changes

The passing of the strongly bipartisan OPEN Government Data Act in 2018, originally introduced by Chairman Kilmer, was a leap forward in making data from the federal government accessible for Americans' personal and commercial use. It created a pathway for agencies to organize and distribute data to other agencies and to the general public in ways that are easy to access and understand. But more must be done. The mandate is largely unfunded, putting the onus on the Chief Data Officers to help educate teams within their respective agencies around the value of disseminating their specific data to the American public.

In 2019, Congress appropriated $8.9B in taxpayer dollars for government statistics. It is the people's data. The people, and those they elect to represent them, should have easy access to the information. There are specific actions Congress must take to continue promoting the use of nonpartisan and unbiased data, including:

- **Increase training for staffers:**
  - Create nonpartisan, data-based onboarding materials for Congressional staffers and help them understand where they can go for trusted data
    - *Anecdotally, our annual report has served as onboarding materials for Congressional staffers looking to better understand government, and we recently had the opportunity to train Congressional staff through an event hosted by the Congressional Management Foundation.*

- **Promote bipartisan discussion about data:**
  - Offer more opportunities within committee meetings and the legislative drafting process for Members of Congress to have bipartisan discussions with government metrics as the foundation
    - *Anecdotally, USAFacts has appreciated the opportunity to present and discuss our annual report on the state of our country by the numbers with Speaker Pelosi and the Democratic Caucus, Minority Leader McCarthy and members of his leadership team, and a bipartisan group of US Senators. We have seen firsthand that Members of Congress are also open to and interested in governing based on a solid understanding of where our country stands based on unbiased data.*

- **Implement measurement around legislation:**
  - Ensure that each piece of legislation has a numeric goal established that can be tracked by entities responsible for implementation and reported back to Congress. Currently legislation is commonly tracked based on dollars spent; however, we recommend measuring impact and effectiveness by establishing a numeric baseline and tracking desired outcomes.

- **Establish more centralized data sources:**
  - ○ Create and fund a centralized data resource for Congress. Congress already has great resources for understanding budget implications of legislation and demographic data (CBO), tax data (JCT), and individual research questions (CRS). However, Congress could establish an additional resource to provide non-partisan non-budgetary data to help Congress understand and set benchmarks for the effects of legislation on the American public. This could be housed within an existing organization such as the GAO.

In closing, Congress and our country's citizens deserve to govern based on trusted data. USAFacts strongly supports your efforts on this committee to continue promoting solutions that increase access, accuracy, and coordination of official data sources.

When trusted data is available to everyone, decision-makers in government and the private sector realize benefits that strengthen the American economy and improve our society. The opportunities for improvement are vast, but the solutions are within our reach and USAFacts remains a strong partner to help achieve them.

Thank you for your time.

+++

Additional Recommendations for the Committee:
Below are 25 key areas where government data can be improved. Each is an example of where data is either missing, incomplete, not timely, or has several competing sources with different information.

1. **Police Use of Force** – We lack complete data on police use of force from local police departments, and aggregates of that data to understand the story nationally. Providing additional support to bring all states into the FBI's National Incident-Based Reporting System (NIBRS) would also help in understanding this area as well. *(Federal, State and Local)*

2. **Improved Public Health Data Sharing** – The COVID-19 pandemic exposed several challenges with how states share data with each other and with the federal government. States reported data inconsistently with differing definitions, standards, time frames, and visualization platforms. Even as the virus rebounds, some states have slowed down COVID reporting or have stopped reporting some metrics altogether. The federal government should mandate and fund emergency and preemptive data collections and provide standards and formats to improve the pipeline of data from local health agencies to the CDC. Converting several underutilized, voluntary surveys like the National Notifiable Disease Surveillance System, into mandatory collections would significantly increase awareness into future health disasters. *(State and Local)*

3. **Timely Income/Spending by Income Group** – More frequently updated data on income, spending, and government transfers by income group are key to understanding people's standard of living and the impact of certain government programs. BEA

currently releases aggregate-only numbers monthly, and IRS numbers lag by a year or more. *(Federal)*

4. **Safety/Reliability of Infrastructure** – Data hasn't updated since 2014 for several key datasets despite the push for infrastructure legislation. Examples: rail transit infrastructure conditions stopped in 2014 and roadway congestion stopped in 2011. *(Federal)*

5. **Cost of Pre-K (to governments and individuals)** – It is extremely difficult to figure out how much preschool costs across the country. Some states release data but there is no aggregation of this other than for Head start. This is relevant to analyzing proposals like those in the American Families Plan. *(Federal/State and Local)*

6. **Healthcare Average Premiums/Out of Pocket Costs** – Data on average premiums is unavailable or extremely hard to come by, even for government programs. MEPS is somewhat good but only for private employer provided healthcare, which likely only covers 20-25% of the population. *(Federal and State)*

7. **Costs of Healthcare** – It is hard to see what is driving health care costs without understanding who gets what medical care, how much it costs, and which insurance program pays. AHRQ is a good start but the cost definitions it uses are very opaque and hard to understand. It could be improved by including more data on out of pocket vs. insurance cost per procedure. *(Federal)*

8. **Gig Workers** – BLS/Census do not provide statistics on the gig economy workforce despite it becoming more prevalent. BLS has some measures on contingent workers but the definitions are out of date for today's economy. An easy win would be adding additional questions to existing survey infrastructure. *(Federal)*

9. **Disaggregation Across Datasets** – Native American populations are often excluded from government datasets. Additional support to the Bureau of Indian Affairs earmarked for increasing sampling in existing surveys would be a huge step in this effort. Additional disaggregation broadly across datasets for the territories would also be beneficial. *(Federal)*

10. **Timeliness of State and Local Spending** – The Census of Governments produces standardized data on state and local spending. However, it lags by two years or more. Having a provisional version of it available earlier would be useful, similar to what census produces for quarterly state tax revenue. *(State and Local)*

11. **NCES Timeliness** – The NCES Digest of Education Statistics is really useful, but most of its data lags 3-4 years. Also, its data is formatted in a way that makes ingestion extremely challenging to export and use. Rethinking this and other statistical compendiums with a modern design that includes more localized breakdowns and machine ingestible formats would make these more useful to modern audiences. *(Federal)*

12. **Unauthorized immigrant population** – The most recent estimate of unauthorized population is from 2018 (and until April of this year, lagged to 2015). Keeping this updated moving forward is important to minimizing misinformation in this area. *(Federal)*

13. **Improved data on the immigrant population residing in the US** – We don't currently have numbers available about who is here and for what purpose that combines people with visas, green cards, and refugee/asylee status. It is hard to see the full picture of

immigration, make decisions on the issue, and minimize misinformation without this data. *(Federal)*

14. **Preliminary / experimental data and associated context across key datasets** – Several agencies have embraced the concept of making their data available in preliminary or experimental format which greatly improves the timeliness of data and enables real-time decision-making. More should do this. *(Federal)*

15. **Continued Funding of the Household Pulse** – The experimental Household Pulse survey started by the Census Bureau during the coronavirus crisis has been a successful asset giving real-time updates on the state of the American people. It should be continued well beyond the pandemic with dedicated funding. *(Federal)*

16. **Improved Data on Outcomes of Various Stages of the Criminal Justice System** – There is a lot of data on the number of people in various stages of the system (arrests, jail, prison, parole), but nothing that effectively explains the flow of people (percent of people who move) from crime committed and/or police interaction to jail to bail to court system to prison to parole to recidivism. *(Federal/State and Local)*

17. **Better Data on the Flow of Students Through the Education System –** Similar to the criminal justice system, we have a lot of information about who is in various parts of the education system, but it is hard to paint a picture of where people drop out and why. Existing systems are state-specific and not comparable across geographies. It would be helpful to have nationally standardized information on who goes to Pre-K, then to K-12, then to college/what type of college, then graduates, then gets a job. *(Federal/State and Local)*

18. **Civic Participation** – The Census Current Population Survey used to ask a number of questions about how people participated in their community and discussed politics, but they have been discontinued. Examples included: participation in school group/neighborhood associations, service organizations, eating dinner with their households, talking to neighbors, contacting a representative, participating in a boycott, etc. *(Federal)*

19. **FOIA Data** – The data made available about FOIA requests themselves is very siloed and challenging to access. It is difficult to assess how many FOIA requests are received and executed and how these requests are categorized (media requests, requests for specific documents, requests for data). *(Federal)*

20. **Open Data Policy Progress** – Understanding data availability and format by agency would be useful in effort to support changes. For example, how much of each agency's data has been made public or can't be (and why), and how much information an agency stores vs. makes accessible to the public vs. is deemed sensitive vs. is yet to be assessed, are important in understanding progress towards better data. PDFs are generally poor formats for releasing usable data and tracking how much is presented this way can help measure progress as well. *(Federal)*