Pursuant to Office of Management and Budget (OMB), Executive Office of the President's Document 83 FR 38183, Pages 38183-38184, Document Number: 2018-16615, the University of Michigan's Economic Growth Institute and Institute for Social Research would like to submit the following response:

The purpose of this RFI response is to speak to **item 8** specifically, regarding the housing of data and its use. The recommendations for a data center are as follows:

1) **A public research university would be the trusted place to house centralized data management and dissemination for GEAR. Universities were established to serve the public good, but have evolved to help the federal government address its critical data management needs and have become a reliable partner for the following reasons:**
    - Mission is for public good and to educate the next generation of data researchers and users
    - Universities are stable entities that exist for centuries
    - Experience with secure management and interpretation of data
    - Experience across the board with data formats and tools, combined with the need to stay abreast of new theories, tools, methodologies and technologies in data management and big data processing with a long-term view
    - Vast resources to explore business, political, social, and economic reasons for how to manage, store, secure, and deploy data
    - History of transparency and innovation
    - Trusted partner of industry, government, and other academic institutions
    - Experience with public/private/government partnerships
    - Demonstration of successful business model development around the use of data for industry and government use

**2) A data management center would best be housed at a university that has a clear understanding and multiple examples of:**

- public/private/government partnerships
- the ability to develop business models using data sets and packages
- attracting multiple revenue sources around data sets
- addressing a wide range of federal government needs, including economic, market, health, and social research
- longitudinal data collection and analysis to serve the public good and inform policies
- public dissemination examples through web based tools, educational forums, and other public engagements

**3) GEAR should look to housing its data management at a university that has:**

- a long-term strategic plan to increase the focus on data management resources
- commitments to big data and processing management
- strong policies to manage data security, integrity
- specifically focused areas of data collection and management, including:
    - Innovation and Economic Development
    - Business and Economics
    - Health Care Cost and Quality
    - Economic Inequality
    - Educational Inequality

The University of Michigan is already at the forefront of collaborations that span organizations, industries, and sectors to create and use data for research to support public benefit. The University of Michigan is also home to many centers considered best practices examples across economic, market, policy, health, and social research that work with practitioners in a variety of fields, but the two (data collection and outreach/application are rarely integrated). We propose a systematic effort to integrate data curation, research, and practice elements of a major university to develop general solutions to problems of pressing federal interest and to learn from application efforts to improve data and future research.

For an anticipated early focus on how Federally owned data could help transform society and grow the economy:

- Are there opportunities for the Federal government to partner with the private sector to improve data architecture/taxonomy, and data quality/hygiene?
    - For decades the university has conducted economic research and analyzed data on behalf of the federal, state and local governments to identify business trends and inform programs that address economic concerns and opportunities. Government-available data is both necessary and in need of translation for real-world economic and business use.
    - Government and private sector data is regularly "pieced together" to try to inform a picture for companies looking to expand their product set, invest in new equipment, acquire a company, launch a spin-off company, or end a product offering with confidence in their business decisions. The university helps companies successfully launch hundreds of new products every year but both the process of collecting the data and the content could be improved, allowing the activity to scale.
    - The University of Michigan Office of Research mission is to "catalyze, support, and safeguard U-M research." Among the states values of the Office of Research are:
        - Altruistic: We derive satisfaction from helping others succeed throughout the research lifecycle, from conceiving paths for exploration to translating discoveries into practice.
        - Collaborative: We promote a culture of collaboration across the U-M research enterprise, building relationships among partners, both internal and external, and recognizing that the sum of what we do is greater than the parts.
        - Ethical: We uphold the highest standards of professional conduct and research integrity.
        - Innovative. We support developing research, scholarship and breakthrough discoveries that solve the grand challenges of today and tomorrow. We seek continuous improvement in our daily operations.
        - Transparent: We embrace the imperative to document and share information in an accessible and timely manner, while respecting our responsibility as stewards of sensitive and confidential information.

- Are there innovative economic models that highlight the value of the data, and would encourage private investment to capture that value both within the Government and across the broader economy? What are the barriers to implementing these models?
    - There is a vast amount of economic, social, and business data that is collected on behalf of governments and used for a singular purpose. The University of Michigan has created partnerships with multiple universities to conduct work and combine data to identify business

needs, technological innovation trends, social and health needs, and economic trends for the public good. The university is also a central source for disseminating information.
    - Census data, as an example, is valuable for several social, governmental, and economic purposes, is often analyzed, and re-packaged multiple ways for business use. There are many commercially available tools and data sets that translate Census data into practical business uses that combine projection methodologies and other non-Census data sets. Similarly the GEAR could identify the commercial needs for data sets, the necessary standardization for some methodologies, and the "packaging" opportunities for industry in a sustainable model.
    - Current barriers are the understanding of how the data is used at the practical, street level implementation to inform business decisions by practitioners and delivery to novices.

- Are there specific data sets that could be further leveraged by the Federal government, start-ups, and the public – that, once scaled, have a significant potential to contribute to the greater good (bolster the economy, improve population health, provide services to the general public, etc.)?
    - Not all government data regarding new technologies is available in an easily digestible format so that businesses - particularly small and medium sized technology and manufacturing businesses - can use to inform decisions. The university assists these companies to understand the availability and shortcomings of data. Data for early stage technical landscape is decentralized and formatted widely, leaving much to be desired in terms of accessibility and usability.
    - Communities looking to build technical economies rely upon government data sources (with the university's assistance) to inform them of necessary strategic planning opportunities. Formats and data available widely vary, and are pieced together to help inform their strategy.

Just a few examples of the breadth and depth of relevant resources that demonstrate a university's data management capabilities, commitment to integrity, and goals for data dissemination:

- **Economic Growth Institute** conducts regular market research to inform manufacturing supply chain and technology businesses each year of new market trends, technical directions, competitive landscapes, and pending regulatory or economic shifts. The Institute also conducts national surveys on best practices in technology development and commercialization.

- **Institute for Research on Innovation and Science** is currently a consortium of more than 30 research universities anchored on an IRB approved data repository housed at the University of Michigan. IRIS was founded in 2015 with support from the Alfred P. Sloan and Ewing Marion Kauffman foundations as well as the contributions of university members to expand the Big 10 UMETRICS pilot into a permanent national resource to support research and reporting to understand, explain, and improve the public value of academic R&D. The institute is governed by member drafted and ratified bylaws and overseen by a board of directors elected by member universities. Universities join IRIS, submit administrative data and contribute $25,000 per year to support the infrastructure necessary for research and reporting. In return, they receive a series of interactive reports describing the economic and workforce impact of their federal research spending, free access to research data for their affiliated researchers, a voice in governance, and the ability to contribute (through a variety of advisory groups) to the development of new reports and data products of potential value. IRIS is organized on a federated core and node model. The core data repository works with organizations called Nodes that are approved by the IRIS

board of directors to access data through IRIS secure data systems for the purposes of improving its value for research and reporting.

Under a partnership with the U.S. Census Bureau's Innovation Measurement Initiative (IMI) project, IRIS transfers data to the Census Bureau where linkages are made to restricted federal data sets that enable statistical analysis and reporting on the post-university careers and earnings of researchers as well as the characteristics of their employers and of vendors that supply goods and services to support federally funded university research.

With the help of researchers at IRIS nodes and the Census Bureau, university data are linked at the individual and grant level to public information on federal grants drawn from Federal Reporter, to PubMed Publications, to US Patents, and to dissertation information shared by ProQuest. Work continues to expand these data linkages with new projects underway to incorporate publication data from Clarivate's Web of Science Dataset, information on federal committee service maintained under the Federal Advisory Committee Act (FACA), detailed information on vendors drawn from Bureau Van Djik's ORBIS data set, the public data release of the Bureau of Labor Statistics Quarterly Census of Employment and Wages (QCEW) and information about software produced in the course of research derived from GitHub repository data.
An annual de-identified data release for research use is constructed from sources for which linkages have been completed and validated. That dataset and its documentation are made available to the research community through a virtual data enclave (VDE) housed at the University of Michigan core facility. A mirror of that dataset along with Title 13 crosswalks to restricted federal data are made available through the Federal Statistical Research Data Center (FSRDC) system. FSRDC research access is managed by the U.S. Census Bureau.

- **The Panel Study of Income Dynamics** (PSID) is the longest running longitudinal household survey in the world. The study began in 1968 with a nationally representative sample of over 18,000 individuals living in 5,000 families in the United States. Information on these individuals and their descendants has been collected continuously, including data covering employment, income, wealth, expenditures, health, marriage, childbearing, child development, philanthropy, education, and numerous other topics. The PSID is directed by faculty at the University of Michigan, and the data are available on this website without cost to researchers and analysts.


The data are used by researchers, policy analysts, and teachers around the globe. Over 4,000 peer-reviewed publications have been based on the PSID. Recognizing the importance of the data, numerous countries have created their own PSID-like studies that now facilitate cross-national comparative research. The National Science Foundation recognized the PSID as one of the 60 most significant advances funded by NSF in its 60 year history.

- **The Michigan Institute for Computational Discovery and Engineering** (MICDE) is the focal point for the wide spectrum of research in computational science and engineering and includes the development and deployment of sophisticated models from nearly every aspect of science and engineering on high performance computers (HPC) to support basic research, product development, and forecasting. Computational Discovery and Engineering (CDE) is an enabling discipline with computation widely accepted as the third mode of scientific discovery on par with theory and physical experimentation.

Research in CDE at U-M is taking place in diverse fields, including cosmology, environmental and climate modeling, population and community ecology, defense and homeland security, nuclear and high-energy device development, materials and nanoscience, biophysics, electrical and electronic design, transportation, geophysical modeling, structural engineering, healthcare engineering, modeling of large-scale engineering systems, information technology, data mining, social network analysis and visualization, and computational linguistics. At MICDE, researchers also define fundamentally new paradigms to blend computer science, future hardware and big data with computational science. Faculty come from many departments within the College of Engineering, the College of Literature, Science and the Arts and beyond.

- **Inter-university Consortium for Political and Social Research** (ICPSR) advances and expands social and behavioral research, acting as a global leader in data stewardship and providing rich data resources and responsive educational opportunities for present and future generations. An international consortium of more than 750 academic institutions and research organizations, Inter-university Consortium for Political and Social Research (ICPSR) provides leadership and training in data access, curation, and methods of analysis for the social science research community. ICPSR:

    - maintains a data archive of more than 250,000 files of research in the social and behavioral sciences. It hosts 21 specialized collections of data in education, aging, criminal justice, substance abuse, terrorism, and other fields.
    - collaborates with a number of funders, including U.S. statistical agencies and foundations, to create thematic data collections and data stewardship and research projects.
    - provides a comprehensive curriculum of intensive courses in research design, statistics, data analysis, and social methodology. ICPSR also leads several initiatives that encourage use of data in teaching, particularly in undergraduate instruction.
    - focuses sponsored research on the emerging challenges of digital curation and data science. ICPSR leads or takes part in many policy initiatives and grant-funded activities that result in publications that address issues related to data stewardship. ICPSR researchers also examine substantive issues related to our collections, with an emphasis on historical demography and the environment.

- **Army Study To Assess Risk and Resilience in Servicemembers** (STARRS) is the largest study of mental health risk and resilience ever conducted among military personnel. Army STARRS investigators looked for factors that help protect a Soldier's mental health and factors that put a Soldier's mental health at risk. Army STARRS ran from July 2009 through June 2015. During that time, research findings were reported to senior Army leadership as the findings became available so the Army had the ability to apply them to ongoing health promotion, risk reduction, and suicide prevention efforts. The goal of extended to contact and reinterview the Army STARRS participants from the All Army Study (AAS), New Soldier Study (NSS), and Pre-Post Deployment Study (PPDS) components of the initial Army STARRS project. In addition to collecting primary survey data for these respondents, Michigan will continue to operate and support users of the Army STARRS Research Data Enclave, which contains the Historical Administrative Data System (HADS) as well as de-identified Army STARRS primary data and Army/DoD data for consenting soldiers.

- **Institute for Social Research** (ISR) is the world's largest academic social science survey and research organization. Established in 1949, ISR has a long history of gathering, managing, and disseminating valuable research data across a wide array of both public and private sector areas. The IRIS local area

network, wide area network, and data infrastructure are managed centrally from within ISR, under the direction of IRIS personnel. The safeguarding and protections of project data are based on a defense-in-depth architecture. Firewalls, secure remote application across platforms, and other boundary controls are implemented based on a risk-based approach that adheres to a least privilege access control model. Specific configuration controls are safeguarded on a strict “need-to-know” basis. Four layers of security maintain access control to data: enclaves, border firewalls and intrusion prevention systems, strong authentication, and configuration/patch management.

- **Michigan Institute for Data Science** (MIDAS)- The University of Michigan (U-M) plans to invest $100 million over the next five years in a new Data Science Initiative (DSI) that will enhance opportunities for student and faculty researchers across the University to tap into the enormous potential of big data. The U-M plans to hire 35 new faculty over the next four years and engage existing faculty across campus, support interdisciplinary data-related research initiatives and foster new methodological approaches to big data, provide new educational opportunities for students pursuing careers in data science, expand U-M’s research computing capacity, and strengthen data management, storage, analytics, and training resources.
  The Data Science Initiative brings together the newly created Michigan Institute for Data Science (MIDAS), Consulting for Statistics, Computing and Analytics Research (CSCAR) and Advanced Research Computing – Technology Services (ARC-TS) to provide a coordinated and comprehensive home for data science as part of Advanced Research Computing (ARC) at the University.
  MIDAS will build an interdisciplinary core faculty of 40 data scientists, and lead programs in Data Science Challenges, Data Science Education and Training, and Industry Engagement.

Submitted by:
Paula Sorrell
Director, Economic Growth Institute
University of Michigan
psorrell@umich.edu
734-998-0148

With input from:
Jason Owen-Smith and Nancy Calvin-Naylor
Institute for Research on Innovation and Science
Institute for Social Research
University of Michigan