To: The Chief Data Officer Council

From: Douglas Hummel-Price, Principal, Public Data Consultancy, LLC[[1]](#footnote-1)

Date: February 23, 2020

RE: Priority Suggestion: CDO Council Task Force to Examine Data Use Cases for Binding Decisions in the Federal Government and Advisory Decisions in the Private Sector

**Executive Summary**

Signed into Law in January 2019, the Title II of the Foundations of Evidence-Based Policymaking Act (“Evidence Act”) mandates a Chief Data Officer (“CDO”) in every federal agency.[[2]](#footnote-2) In addition to wide-ranging authority regarding data governance within their respective agency, each CDO is charged with participating in the newly formed Chief Data Officer Council at the Office of Management and Budget.[[3]](#footnote-3) The legislation provides significant leeway for OMB and the CDO Council to decide how to achieve the goals set forth in the Evidence Act. **The Council should make it a priority to establish a Data Use Task Force designed to evaluate the ethics of specific use cases of agency data for inter- and intra-agency research as well as private sector use of government data.**

**Background**

The fundamental impetus for the Evidence Act comes from a combination of the exponential increase in data collection and use by both the public and private sectors, a perceived lack of unified vision by federal agencies regarding data governance, and an understanding that prior to the act, incentive structures did not encourage agencies to work together or share data easily between them. Researchers have grown increasingly aware of the necessity of a solid ethical framework around data science.[[4]](#footnote-4) Data ethics are rife with trade-offs; these trade-offs are explained well in a memo by the Information Commissioner’s Office, the UK’s “independent authority set up to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals.”[[5]](#footnote-5) The ICO statement explains that there are nearly always trade-offs between combinations of the following data ethics concepts: privacy, accuracy, fairness, explainability, and security.

The proposed Data Use Task Force (“Task Force”) would primarily focus on privacy, accuracy, and fairness. Data security, while incredibly important, does not fall under the purview of the Council. The US is lacking in comprehensive data protection legislation, but under the current patchwork of laws, the Federal Trade Commission is broadly empowered to enforce federal data protection and privacy regulations.[[6]](#footnote-6) Explainability *does* fall under the responsibilities of the Council, but we believe that it deserves its own task force, outside the scope of this recommendation. **Therefore, the primary lenses through which the Task Force would examine data use cases are privacy, accuracy, and fairness.**

**Legal Grounding**

The primary function of the Task Force is *ethical*, not legal, analysis. Law and ethics are not synonymous; in fact, a basic function of ethical analysis is looking beyond legal considerations to more fundamental concerns.[[7]](#footnote-7) However, before we discuss how to incorporate the primary lenses above in ethical analysis, we must first establish the legal foundation for the Task Force. The Council has five codified purposes and functions laid out in the Evidence Act.[[8]](#footnote-8) Of these, the first, second, and fourth form the basis for the recommendation to establish a Task Force:

§3520A. Chief Data Officer Council

(b) Purpose and Functions.-The Council shall-

(1) establish Governmentwide best practices for the use, protection, dissemination, and generation of data;

(2) promote and encourage data sharing agreements between agencies;

(4) consult with the public and engage with private users of Government data and other stakeholders on how to improve access to data assets of the Federal Government;

The Task Force should examine three broad categories of use cases for data, each corresponding with one of the functions: intra-agency, inter-agency, and private sector. Purpose One provides the basis for intra-agency data use and dissemination. Purpose Two provides the basis for inter-agency data sharing, and Purpose Four establishes cause for the Task Force to look at private sector use of federal data.

**Task Force Structure: Appointing Membership**

As of late 2019, the exact logistics and structure of the Chief Data Officer Council are yet to be announced.[[9]](#footnote-9) This provides both opportunities and challenges for our recommendation. Without a clear understanding of the specific structure and meeting plans for the Council, our recommendation is necessarily in general terms. However, the blank slate allows us significant freedom to present how we believe the Task Force should play out. The Council should begin by selecting seven Council members to serve on the Task Force for one-year terms.[[10]](#footnote-10) These members should all be Chief Data Officers of their respective agencies, rather than the handful of pro forma and OMB representatives also on the Council.

In a given term, the seven members should first pick one among them to act as chair; the chair’s purpose is primarily logistical, and their opinions on substantive matters before the Task Force should not carry additional weight as chair. The purpose of one-year terms is several-fold. First, the type of dedicated analyses described below requires careful consideration of details in a variety of contexts. A bit of analysis fatigue is to be expected after twelve months on the Task Force, so rotating in new members will keep the analyses fresh. Second, a foundational question in data ethics is “Whose ethics?” Both demographic and ideological diversity are vital to proper ethical analysis of data use cases.[[11]](#footnote-11)

A rotating membership increases the number of perspectives, in turn decreasing the likelihood of blind spots in Task Force analyses. Third, a Task Force with static membership runs the risk of something analogous to regulatory capture both in the traditional private sector sense and in the sense of any one agency having undue influence.[[12]](#footnote-12)

**Task Force Structure: Use Case Nominations and Evaluations**

Before analyzing any given data use case, the Task Force must first establish procedures for nomination and selection of use cases to examine. Initially, a list of potential use cases should be compiled; we recommend three channels of nomination. Most obviously, members of the CDO council can submit cases. Secondly, any government employee can submit cases that directly regard their agency. This should not take the place of any whistleblower or inspector general processes already in place.[[13]](#footnote-13) Lastly, the Task Force should create a public submission channel, perhaps housed on Data.gov.[[14]](#footnote-14) We will leave it up to the Task Force how to prioritize the nominations, but the general motivating premise should be looking at a variety of use cases, with a preference on specific cases that result in analyses that can be generalized to a broader set of cases.

**Analysis Framework: Three Lenses**

With the logistical and legal housekeeping covered, we now discuss the framework within which the Task Force should consider each use case. We recommend that the analyses focus on the three lenses mentioned above: privacy, accuracy, and fairness. The general workflow is in Figure 1 below:

**Federal Employees**

**General Public**

**CDOs**

**Use Case Nominations**

**Domain**

**Experts**

**Privacy Lens**

**Preliminary Phase**

**Recommendations Phase**

**Accuracy Lens**

**Technical Experts**

****

**Fairness Lens**

**Figure 1: Task Force Workflow**

Once a case is selected for review, the Task Force enters the preliminary phase. This phase consists of general fact-finding. The Task Force should consult with both technical and domain experts to gather as much information as it can in the time allotted. This phase should be approximately a month to give experts time to weigh in. Once the relevant facts have been gathered, the Task Force then evaluates each of the three analysis lenses as outlined below.

**Analysis Lens: Privacy**

The general public professes to place great import on privacy.[[15]](#footnote-15) Therefore, the Task Force should enumerate groups of stakeholders potentially impacted by privacy concerns. Generally, the individuals whose data are being used will form the largest group of stakeholders. The Task Force should consider the ramifications of various types of privacy infringements. De-identifying and anonymization should not be considered trump cards to privacy concerns. Research has shown that these concepts are not as effective as they profess to be and that most attempts at anonymization are merely de-identification with re-identification always a possibility.[[16]](#footnote-16) As a worst-case scenario, the analysis should consider the potential harm should the data set in question be completely re-identified and released to the public. For example, the mass release of NIH health data could have severe impact on employment and insurance prospects as well as social stigma. On the opposite end of the spectrum, consider a data set that contains information that is already public; a leak of this data set would have minimal impact on the individuals represented in the data set.

While individual privacy concerns are relatively straight-forward, the Task Force should also consider group privacy concerns. For example, crime data about young black men might contribute to stereotypes that negatively impact members of that group, even if they are not in the specific data set. Technical experts can help the Task Force identify potential avenues of inadvertent and nefarious violations of privacy, whereas domain experts can inform the real-world consequences of such violations. Ultimately, the goal here is amassing a list of potential concerns to guide how to best use the data in the question as well as whether such use is advisable in the first place.

**Analysis Lens: Accuracy**

How accurate is the data? Inaccuracies can creep into federal data sets in both practical and theoretical ways. On a practical front, inaccuracies can result from manual data entry errors or similar “technical” failures. In discussing and establishing best practices for data governance, the Council will discuss ways to minimize these inaccuracies moving forward. However, in the short-term, some data sets will be more vulnerable to practical inaccuracies than others. Theoretical challenges also contribute to inaccuracies as a variable may not be an accurate proxy for what it claims to measure. Educational testing presents this challenge – standardized tests in particular are known to be inaccurate measures of underlying aptitude.[[17]](#footnote-17) At this stage, the analysis should be focused on factual discussions of accuracy; prescriptive discussions will occur in the recommendations phase.

The Task Force should also consider the accuracy of the proposed methodology’s outputs. Is the model being used to decide who should be kept in prison pre-trial, for example? In such a case, a false positive (being denied bail when they are not a risk) unduly strips individuals of their liberty, while a false negative (being released and then reoffending or fleeing) potentially results in violence or injustice. It can be difficult to predict even in broad strokes how accurate a given model will turn out, so the focus here should be on creating criteria upon which to evaluate the model if it is eventually implemented.

**Analysis Lens: Fairness**

The third lens to consider is fairness. Will some groups be more impacted than others, positively or negatively? Does the instance take advantage of federal data to give an unfair leg up to a private organization? What is the intended goal of the use case? For these and other questions of fairness, domain experts are vital.[[18]](#footnote-18) The Task Force should consider the input of non-profit groups that represent traditionally under-represented groups. This is also the prime place for the Task Force to attempt to illuminate its blind spots. Seven members will never cover every possible group, and they should be wary of assuming that a single member can represent all concerns of the group they’re a part of – e.g., a fifty-year old Hispanic gay man does not represent all individuals who are middle-aged, Hispanic, gay or male. Non-profits will likely jump at the opportunity to have their voices heard, and of the three lenses, fairness benefits the most from these outside perspectives.

**Recommendation Phase: Three Recommendation Types**

Once all three lenses have been evaluated for a given case, the Task Force can move to the recommendation phase. The recommendations will not be legally-binding, but the very structure of the CDO Council provides them weight. Each recommendation should be posted publicly on data.gov. Should the agency ultimately disagree with the recommendation, the CDO of that agency should be required by the Council to submit a statement explaining so, similar to an “Against Medical Advice” form in a medical setting. This contributes to the relevance of the recommendation without a legal mandate that could make things more complicated for the Task Force. The recommendation should be drafted and voted upon before presented to entire Council, who can only reject the recommendation with a two-thirds majority vote.[[19]](#footnote-19) Generally, the recommendations will fall into one of three categories:

1. Minimal Concerns
2. Move Forward with Caution
3. Reconsider Use

**Recommendation Type One: Minimal Concerns**

The first recommendation type represents situations where there is minimal risk of privacy, accuracy, and fairness concerns. The Council should do all it can to facilitate the use case. Since the relevant agencies will have representatives on the Council itself, the Task Force’s recommendation should de-facto facilitate the use in question.

**Recommendation Type Two: Move Forward with Caution**

The second recommendation type represents situations in which a use case is feasible but presents risks. The outputs from the three lenses of analyses should highlight what needs to be done to minimize risk and maximize benefit. The Council should offer whatever resources it has available, particularly best practices and tools, to facilitate the use case. The Task Force should be sure to provide metrics by which the relevant agency can check itself as it implements the use case. In the future, the Task Force or Council may want to develop a procedure to check back in with agencies mid- and post-implementation.

**Recommendation Type Three: Reconsider Use**

The final recommendation type applies to situations in which the Task Force has grave reservations about the use case. Whether the privacy risk is too high, the accuracy of the data or model too suspect, or the impact of use too unfair to sub-populations, the Task Force is effectively stating that the proposed use case is unethical. The Council will not provide any additional support to the agency in question should it decide to move forward despite the recommendation. Further, the CDO of the agency must state in clear, written terms that the agency understands the recommendation and risks associated with the use case. These statements should be posted to Data.gov in an effort to promote transparency and provide citizens with the information they need to be civically engaged.

**Conclusion**

The bi-partisan passage of the Evidence Act indicates that data governance is not the provenance of either major party. The Chief Data Officer Council presents a great opportunity for the federal agencies to establish government-wide, standard ethics regarding use cases of federal data. The formation of a Data Use Task Force within the Council will provide consistency and clarity in setting these standards. **We recommend that the Council prioritize creation of such a Task Force.**

1. The Public Data Consultancy, founded in the fall of 2018, is comprised of a small team of data scientists from the public policy sphere. Mission Statement: “The Public Data Consultancy shall use its expertise to perform research into the ethical use of public data at the federal, state, and local levels and produce recommendations to relevant government bodies.” [↑](#footnote-ref-1)
2. Paul D Ryan, “Text - H.R.4174 - 115th Congress (2017-2018): Foundations for Evidence-Based Policymaking Act of 2018,” Congress.gov, January 14, 2019, https://www.congress.gov/bill/115th-congress/house-bill/4174/text#H38F2A921F44747F19758F57600557AC4) [↑](#footnote-ref-2)
3. “Stakeholder Engagement in Developing the Federal Data Strategy and 2020 Action Plan”, The Office of Management and Budget, December 2019. https://strategy.data.gov/assets/docs/2019-2020-federal-data-strategy-revisions-based-on-feedback.pdf. [↑](#footnote-ref-3)
4. Mark Latonero and Zachary Gold, “Data, Human Rights & Human Security,” DataSociety.net, June 22, 2015, https://datasociety.net/pubs/dhr/Data-HumanRights-primer2015.pdf) [↑](#footnote-ref-4)
5. Reuben Binns and Valeria Gallo, “Trade-Offs,” Information Commissioner's Office, July 25, 2019, https://ico.org.uk/about-the-ico/news-and-events/ai-blog-trade-offs/)https://ico.org.uk/about-the-ico/news-and-events/ai-blog-trade-offs/. [↑](#footnote-ref-5)
6. “Data Protection 2019: Laws and Regulations: USA: ICLG,” International Comparative Legal Guides International Business Reports (Global Legal Group, July 3, 2019), https://iclg.com/practice-areas/data-protection-laws-and-regulations/usa) [↑](#footnote-ref-6)
7. To illustrate this, consider free speech as applied to neo-Nazis. Legally, they have fairly broad rights to speak their views as protected by the first amendment, but ethically, their behavior is repugnant. [↑](#footnote-ref-7)
8. Paul D Ryan, “§3520A. Chief Data Officer Council” [↑](#footnote-ref-8)
9. Brandi Vincent, “Federal CIO: Chief Data Officer Council Is Coming Very Soon,” Nextgov.com (Nextgov, November 5, 2019), https://www.nextgov.com/cio-briefing/2019/11/federal-cio-chief-data-officer-council-coming-very-soo/161098/) [↑](#footnote-ref-9)
10. We leave it up to the Council how to specifically do this. One path could be soliciting volunteers, and then having a ranked vote where each member ranks their top seven choices. Another path could random selection. [↑](#footnote-ref-10)
11. Sophie Searcy, “Why Data Scientists Should Make a Commitment to Diversity,” InformationWeek (Information Week, December 7, 2018), https://www.informationweek.com/big-data/big-data-analytics/why-data-scientists-should-make-a-commitment-to-diversity/a/d-id/1333403) [↑](#footnote-ref-11)
12. John Taylor, “The Danger of Regulatory Capture,” PolicyEd, October 11, 2018, https://www.policyed.org/intellections/danger-regulatory-capture/video) [↑](#footnote-ref-12)
13. NB: Nothing done by the Task Force should be considered to justify any action deemed by the courts or Inspectors General illegal or improper. [↑](#footnote-ref-13)
14. The Evidence Act contains a number of provisions regarding Data.gov. A public submission channel housed there would contribute to cohesiveness of federal data policy, one of the main goals of the Evidence Act. [↑](#footnote-ref-14)
15. There is reasonable debate to be had whether the public’s behavior matches their professed sentiment, but at least when asked, the public cares a lot about privacy. “EPIC - Public Opinion on Privacy,” Electronic Privacy Information Center, accessed February 23, 2020, https://epic.org/privacy/survey/) [↑](#footnote-ref-15)
16. Arvind Narayanan and Vitaly Shmatikov, “Myths and fallacies of ‘Personally Identifiable Information.’” Commun. ACM 53, 6 (June 2010), 24–26. DOI: https://doi.org/10.1145/1743546.1743558 [↑](#footnote-ref-16)
17. Peter Simmons , “Predictors of College Success Memo,” Unpublished Internal Memo, The Center for Research on Children in the United States (CROCUS), September 28, 2019. [↑](#footnote-ref-17)
18. Moritz Hardt, “How Big Data Is Unfair,” Medium, September 126, 2014, https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de) [↑](#footnote-ref-18)
19. The specific threshold is somewhat arbitrary. The Council will need to discuss the best way to ensure that the Task Force is respected without opportunity for a small number of people to rock the boat unnecessarily. Some sort of veto system could also be devised. [↑](#footnote-ref-19)