**June 2019**

**Social and Decision Analytics Division**

**Data Science Ethical Protocol Checklist**

**Why is a Checklist for Data Science Research Ethics Needed?**

As part of our research process, we check to ensure that our projects are meeting ethical requirements throughout the life of the project. In addition to the Institutional Review Board (IRB) review, checklists act as an ***aide****-****mémoire* to good research practice**. Our ethics checklist is intended to support ethical considerations throughout a project. The checklist prompts the making of clear statements of intent, mechanisms of approach, and consideration of hazard arising from research in a manner which can be understood by the public and research professionals alike. While some of the items appear to be beyond the scope of ethics alone, any matter that may affect the success of research is of indirect ethical interest if it may expose respondents or researchers to exploitation or risk. Our checklist has been designed as a starting guide for researchers. ***Even with the checklist, an ethical review is complex and will involve discussions with the project team and beyond.***

**How to use the Ethics Checklist Template for Data Science Projects**

It should be used as a self-check by individual researchers and research teams at the project initiation, mid-term review, and final review. It should also be used anytime during the project lifecycle as ethical questions come up. The ethics checklist also becomes an important component of the overall research project documentation.

**How to use Ethics Checklist Template for Peer Review of Data Science Projects**

The checklist should be included in the peer-review of the research project. Specifically, one or more external (to the team) researchers should review and provide feedback on the checklist at project initiation and when substantial adjustments to the checklist are made during the life of the project.

Reviewers looking at your project will require:

1. A completed SDAD Ethics checklist
2. Copies of consent forms or scripts (if applicable)
3. A copy of the research proposal
4. Research findings at time of review

**Social and Decision Analytics Division**

**Data Science Framework**

Before presenting the Ethics Checklist, it is useful to review our Data Science Framework that we use to operationalize our data science research. The Framework provide the structure for executing transparent and repeatable data science research.

A critical component is the Ethics Review that has a defined link to every step in the Framework. The Ethics Checklist is organized around these research steps.



**Data Science Framework**. The data science process starts with the research question and continues through the following steps: data discovery, inventory, screening, and acquisition; data quality assessment (data profiling); data preparation and linkage; data exploration; assessment of the fitness-for-use; and statistical modeling and data analyses. **Social and data ethics are addressed and integrated throughout the life of the project.**

**References**

Keller, S., Lancaster, V., & Shipp, S. (2017). Building capacity for data-driven governance: Creating a new foundation for democracy. *Statistics and Public Policy*, *4*(1), 1-11.

Keller, S., Shipp, S., Korkmaz, G., Molfino, E., Goldstein, J., Lancaster, V., Pires, B., Higdon, D., Chen, D. & Schroeder, A. (2018). Harnessing the power of data to support community‐based research. *Wiley Interdisciplinary Reviews: Computational Statistics*, *10*(3), e1426.

**Social & Decision Analytics Division**

**Ethics Checklist Template for Data Science Projects**

**Project Initiation -** *Recognize and affirm that all project plans will incorporate regular checks, discussion, and documentation to ensure adherence to the ethical principles of research.*

**Problem Identification (Relevant Theories and Working Hypotheses) -** *Establish the ethical basis for undertaking the project as well as the project requirements of both the protection of research participant and the equitable allocation of all potential project benefits and risks.*

1. What are the expected benefits of the project to the “public good,” and do they outweigh potential risks to participant welfare?
2. Are there implicit assumptions and biases in the framing of the project regarding the studied communities and how will they be addressed*?*
3. What type of Institutional Review Board (IRB) approval process is needed? Has the team reviewed the IRB protocol? If an IRB is not required, why not?

**Data Discovery, Inventory, Screening, & Acquisition -** *Consider potential biases that may be introduced through the choice of datasets and variables.*

1. Do the data include disproportionate coverage of the different communities of study?
2. Do data have adequate geographic coverage?
3. Have checks and balances been established to identify and address implicit biases in the data?

**Data Ingestion and Governance -** *Put in place data platforms and processes to ensure data transfer, storage, and database development adheres to data governance agreements and best practices for data quality assurance.*

1. Have team members reviewed standard operating procedures (SOPs) and data management plans?
2. Do additional procedures need to be defined for this project?

**Statistical Modeling & Analysis -** *Establish transparency in methods, results and limitations.*

1. Have project methods and outputs been made as transparent as possible?
2. Are the potential limitations of the research clearly presented?
3. Should the research be used as the basis for policy action, have the predicted benefits and social costs to all potentially affected communities been considered?

**Fitness-for-Use Assessment -** *Critically asses the overall utility of the results in achieving the predicted benefits of the study, to be transparent about potential limitations of the study, and to ensure that unintended biases haven’t been introduced as a result of data choice and model refinement.*

1. What are the limitations of the results? Are the results useful given the purpose of the study?
2. Do the statistical results support the potential benefits of the study previously stated?
3. Do the statistical results support the mitigation of the potential risks of the study previously stated?
4. Have any data been deemed unusable that require revisiting the question of potential biases being introduced through the choice of datasets and variables (from the data discovery to fitness-for-use phase)?

**After project debriefing-** *Summarize questions and actions taken to reinforce the process of ethical consideration on all continuing and future projects.. Establish protocols for replication and expansion of the research findings, and information dissemination.*

1. Did key ethical questions arise during the research and, if so, how were they addressed? How could they be addressed differently on future projects?
2. Are research protocols, methods and data available to other researchers? If so, in what way, and, if not, what factors are limiting the ability to do so?