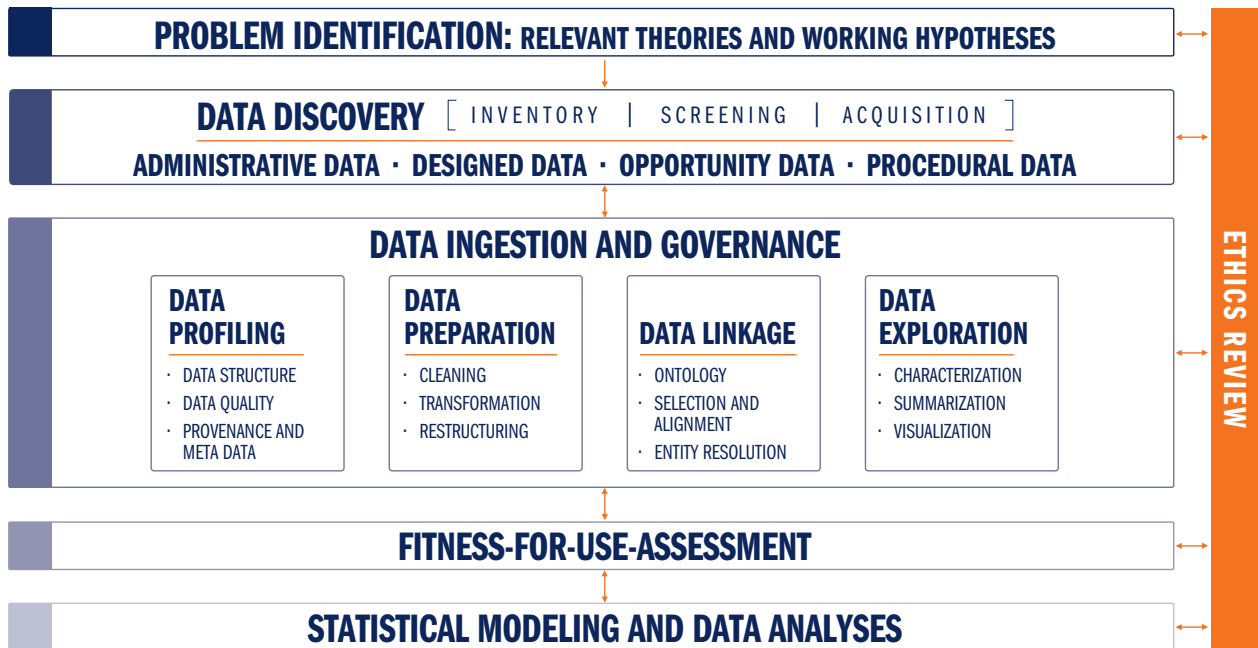


## Social and Decision Analytics Division Data Science Framework



The Data Science Framework 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 (Keller, Shipp, Korkmaz, et al, 2018, Keller, Lancaster, & Shipp, 2017).

### The SDAD Ethics checklist aligns with our Data Science Framework (see page 2)

### 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 Project

### 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?

### Data Discovery, Inventory, Screening, & Acquisition

*Consider potential biases that may be introduced through the choice of datasets and variables.*

4. Do the data include disproportionate coverage of the different communities of study?
5. Do data have adequate geographic coverage?
6. 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.*

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

### Statistical Modeling & Analysis

*Establish transparency in methods, results and limitations.:*

9. Have project methods and outputs been made as transparent as possible?
10. Are the potential limitations of the research clearly presented?
11. 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 assess 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.*

12. What are the limitations of the results? Are the results useful given the purpose of the study?
13. Do the statistical results support the potential benefits of the study previously stated?
14. Do the statistical results support the mitigation of the potential risks of the study previously stated?
15. 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.*

16. Did key ethical questions arise during the research and, if so, how were they addressed? How could they be addressed differently on future projects?
17. 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?