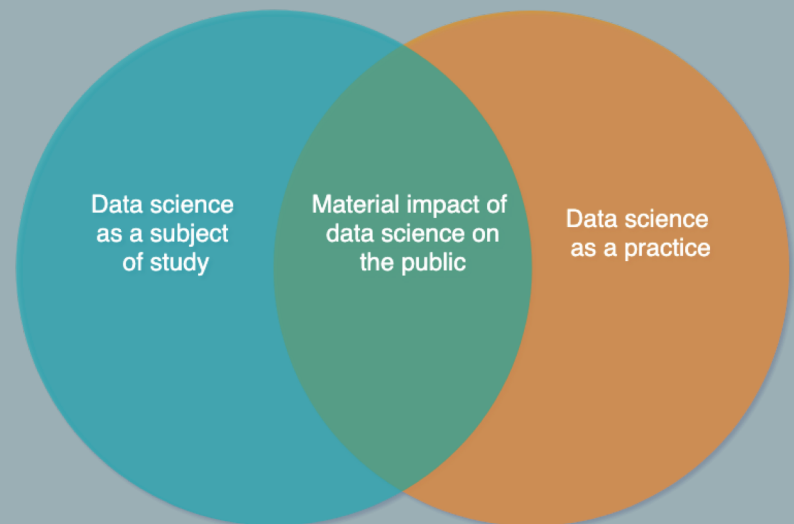# FROM DATUM TO DATA: A QUALITATIVE RESEARCH PROTOCOL FOR DATA SCIENCE IN THE PUBLIC SECTOR

William A. Jenkins, M.A., M.L.I.S.
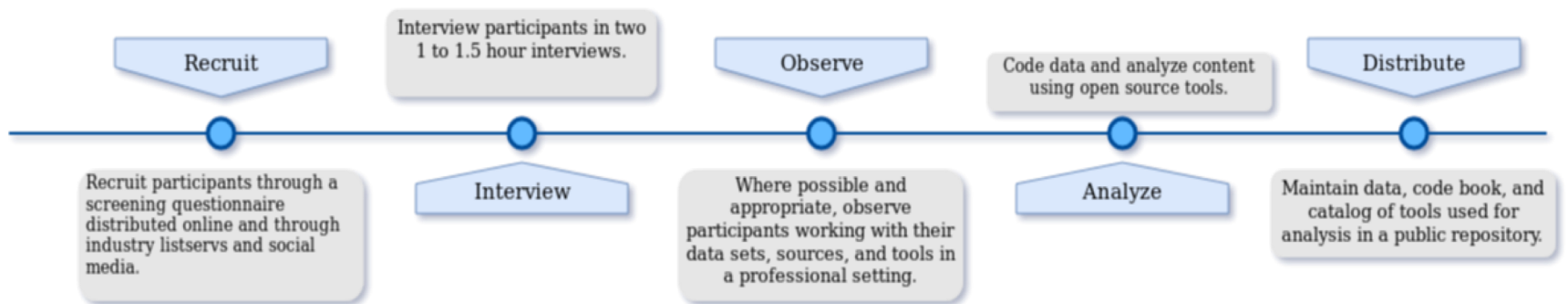Sponsor: University of Washington/Cascadia Urban Analytics Cooperative

BACKGROUND:

Data science is visibly changing the material experience of the public, but data scientists working in the public sector and their practices are an as-yet-understudied population. Studying this population and their emergent practices requires a transparent, reproducible research design that can serve as the start of a toolkit for future sociotechnical research.

Data science as a subject of study — Material impact of data science on the public — Data science as a practice

SOLUTION:

A research protocol for preregistering interpretive qualitative research into data science in the public sector that emphasizes transparency and reproducibility.

Recruit — Recruit participants through a screening questionnaire distributed online and through industry listservs and social media.

Interview — Interview participants in two 1 to 1.5 hour interviews.

Observe — Where possible and appropriate, observe participants working with their data sets, sources, and tools in a professional setting.

Analyze — Code data and analyze content using open source tools.

Distribute — Maintain data, code book, and catalog of tools used for analysis in a public repository.

RESULTS:

A publicly available GitHub repository for researchers distributing:

- Qualitative research protocol
- Methodology statement
- Example code book

"Data becomes a medium for thinking about items, observations, actions, or people – from datum to data… context matters to the understanding of data, and data themselves serve as a media for negotiating meaning."
– Gina Neff, et. al. "Critique and Contribute: A Practice-Based Framework for Improving Critical Data Studies and Data Science" (2017)