APPENDIX A, Data Sharing Plan

All unrestricted data generated during the proposed project life cycle will be documented in accordance with conventionally accepted standards and deposited to a publicly available venue.

1. Quality-controlled raw data as well as processed data used in publications will be made available.
2. Workflows, methods, and procedures will be described exactly and documented carefully for precise reproduction of results from raw data.
3. Primary Investigators (PIs) and Co-PIs will disseminate results from the research through presentations at public lectures; scientific institutions and meetings; and/or publications in major journals.
4. Software and its documentation will be made freely available through an online source code repository.
5. Unrestricted data and associated information generated over the proposed project life cycle will tentatively be made publicly available before November 1, 2028.
6. Research data and associated information restricted by law, regulation, policy, or security conditions will be excluded.

PI Data-Sharing Experience

The Open-Source sYSTem for Entity Resolution (OYSTER) is a project of the Center for Advanced Research in Entity Resolution and Information Quality (ERIQ) at the University of Arkansas at Little Rock. The project was started in 2008 and has been ongoing to the present time. Initially hosted on SourceForge.net, the project (oysterer) was migrated to BitBucket.org https://bitbucket.org/oysterer/oyster/src/master/ in 2012. The project has been downloaded 8,639 times from SourceForge.net. Unfortunately, BitBucket does not provide download statistics.

There are several documents and publications available describing OYSTER functionality including a Reference Guide on the BitBucket download site, descriptions in two books (Talburt J. , 2011), (Talburt & Zhou, Entity Information Life Cycle for Big Data, 2015), a book chapter (Talburt & Zhou, A Practical Guide to Entity Resolution with OYSTER, 2013), and conference papers (Kobayashi, Eram, & Talburt, 2018), (Talburt & Zhou, 2012), (Nelson & Talburt, 2011), (Zhou, Talburt, Su, & Yin, 2010).