

International Journal of Population Data Science

Journal Website: www.ijpds.org



Building a repository for record linkage

Leonard, SH^{1*}, Israel, A¹, Levenstein, M¹, and Alexander T¹

¹ICPSR/University of Michigan

ICPSR is building LinkageLibrary, a repository and community space for researchers involved in linking and combining datasets, as a collaboration between social, statistical, and computer scientists. Unlike surveys or experiments where causal and outcome variables are measured in tandem, it is often necessary when working with organic, non-design data to link to other measures. This makes linkage methodologies particularly important when conducting analyses using administrative data. A common benchmarking repository of linkage methodologies will propel the field to the next level of rigor by facilitating comparison of different algorithms, understanding which types of algorithms work best under different conditions and problem domains, promoting transparency and replicability of research, and encouraging proper citation of methodological contributions and their resulting datasets. It will bring together the diverse scholarly communities (e.g., computer scientists, statisticians, and social, behavioral, economic, and health (SBEH) scientists) who are currently addressing these challenges in disparate ways that do not build on one another's work. Improving linkage methodologies is critical to the production of representative samples, and thus to unbiased estimates of a wide variety of social and economic phenomena. The repository will accelerate the development of new record linkage algorithms and evaluation methods, improve the reproducibility of analyses conducted on integrated data, allow comparisons on same and different data, and move forward the provision of privacy-aware integrated data. The presentation will focus on lessons learned while building the repository and the community, and introduce the LinkageLibrary website.

*Corresponding Author:

Email Address: hautanie@umich.edu (SH Leonard)

