

International Journal of Population Data Science

Journal Website: www.ijpds.org



Assessing Linkage Quality: Metrics, Insights, and a Reusable Python Toolkit

Michael Edwards¹, Anahita Zamani¹, Sean Scully¹, Roshan Toby¹, and Simon Thompson¹

¹SeRP, Swansea, United Kingdom; Swansea University, Swansea, United Kingdom

The increasing use of linked datasets for research and administrative purposes has highlighted the need for robust quality metrics and insights to support their trusted use. Any linkage pipeline inherently contains assumptions and biases in its input data and models, which influence the links it produces and can affect resulting analysis. Commonly used metrics such as match rates, global precision and recall, and F-scores, offer value but are often debated in terms of their effectiveness and interpretation. Developing high-quality linkage models typically requires more granular insights, including qualitative assessments and holistic descriptors of linkage outputs to guide iterative refinement by linkage practitioners. To support consistent, effective, and reusable quality reporting, we developed a Python package that provides a suite of metrics for linkage evaluation. The package includes measures of edge and cluster quality and supports quality analysis within a single linkage model or across multiple models as needed. Built on DuckDB, it evaluates identified record pairs through edge-level metrics, examines grouped records via cluster-level metrics, and offers utility tools for exploratory quality assessment. The metrics are defined as SQL-like query strings, enabling flexible integration into existing linkage workflows or conversion to alternative syntaxes. Delivering clear insights into linkage models and their outputs improves not only link quality but also the confident and appropriate use of linked datasets. As the development of metrics and quality-assessment approaches continues, we welcome contributions from both data linkers and data users to help define measures that are interpretable, informative, and actionable.

