International Journal of Population Data Science

Journal Website: www.ijpds.org

Operationalising Graph-Based Linkage with Directory-Structured Project Frameworks

Michael Edwards[1], Anahita Zamani[1], Roshan Toby[1], Sean Scully[1], and Simon Thompson[1]

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

Motivations Consistently structuring complex linkage projects is essential for transparency, reproducibility, and effective management, and should ideally align with recurring linkage workflow tasks. We present a method that translates an abstract graph-based linkage workflow into a directory-centric representation, mapping subgraphs and composite graphs directly onto a file system. Record linkages are materialised as discrete subgraphs stored within dedicated directories containing predicted edges, metadata, quality metrics, and version history, which can be used in subsequent linkages. Composite graphs are generated by systematically aggregating edge lists followed by a clustering step, with the directory tree enabling selective inclusion of edge sets in resulting graphs. This structure establishes direct correspondence between conceptual graph units and storage locations, enabling transparent navigation from high-level composite cohorts to constituent subgraphs. Versioned directories preserve graph lineage, support provenance tracking and comparison of linkage iterations, and allow reproducible graph reconstruction. The separation between subgraph and composite graph layers supports modular updates, with subgraph revisions propagating to dependent composite structures without destabilising the broader system. By mapping graph concepts to concrete filesystem constructs, the approach reduces cognitive overhead for practitioners, operationalises edge provenance, and provides an auditable, reproducible foundation for large-scale linkage workflows. The approach abstracts across data domains, entity types, and linkage techniques, offering a scalable mechanism for managing graph complexity while maintaining control, transparency, and governance in real-world data linkage environments. Early application demonstrates improvements in linkage management, quality inspection, and reproducibility, illustrating the practical value of combining conceptual graph structures with disciplined project organisation.

https://doi.org/10.23889/ijpds.v11i5.3549