

Case Study: Profiling, Cleansing, Transformation of Event & Identity Data

Project Scope	Support the implementation of a new core business system for a state government department by undertaking the profiling, cleansing, transformation and migration of the legacy data.
Value	\$2M+
Requirements	Profile, analyse, cleanse, transform and migrate data from a core legacy system. The legacy data comprise over 60 million records, representing 200 years of information, held in a proprietary flat file format. Following profiling and cleansing, legacy data were transformed to a new structure and migrated to an RDBMS.
Approach	<ul style="list-style-type: none"> • Profile legacy data relating to persons / events by applying approximately 34 billion validations to identify data issues for manual and automated remediation. • Cleanse and standardise data according to the business rules. Applied approximately 13 billion cleansing rules as part of this process. • Extract person data from events and create person records. • De-duplicate persons and associate unique people back to events in line with the person-centric data model we developed. • Map records from the legacy system's flat file structure to the new relational schema and undertake transformations as necessary. • Name and address parsing and address matching. • Migrate cleansed / transformed data to a staging database, from which it can be imported into the new system. • Maintain an audit trail and report on the results of the cleansing and migration activities.
Outcome	<p>Created a person-centric relational database to replace the legacy event-based structures.</p> <p>Undertook multiple cycles of profiling, cleansing and transformation to reduce initial 'error' count from 300 million items to only a few thousand that required reference to physical documents.</p>
Business Benefits	Provided business assurance of accurate and enhanced records of state significance, ready for migration to the new business system.