

ABSTRAK

Pada perkembangan teknologi saat ini, data memiliki peran penting untuk mendukung pencapaian tujuan bagi perusahaan. Pentingnya data bagi perusahaan yaitu untuk pemenuhan kualitas menunjang kebutuhan bisnis perusahaan. Kualitas tinggi yang dimiliki oleh data bernilai penting bagi perusahaan. Namun banyak terjadi kesalahan pada data sehingga mengurangi kualitasnya. Data dengan kualitas yang rendah yaitu data tersebut tidak akurat, tidak lengkap, atau kadaluwarsa. Perlu adanya manajemen kualitas data atau *Data Quality Management* untuk mengatur peningkatan kualitas data menjadi data yang konsisten, akurat, lengkap, tepat waktu, dan unik. Mengatur kualitas data tersebut perlu adanya *data cleansing*. *Data cleansing* adalah metode untuk memperbaiki data-data yang berkualitas rendah tersebut dengan menghasilkan data dengan kualitas tinggi. Oleh karena itu penelitian ini akan membahas analisa dan perancangan dekomposisi pemecahan paket-paket proses untuk *data cleansing* guna memperbaiki kualitas pada data yang tidak sesuai kebutuhan perusahaan. Perancangan penelitian ini menggunakan metode *iterative incremental*, dimana *iterative incremental* merupakan metode dalam pengembangan yang terdiri dari aktivitas-aktivitas yang disusun secara berurutan dan bersifat fleksibel dalam perubahan yang akan terjadi dalam proses pengembangan. Pada perancangan yang dilakukan, penulis memberikan beberapa solusi berdasarkan analisa untuk memenuhi kebutuhan dalam proses *data cleansing* bagi perusahaan secara fleksibel dengan dekomposisi paket proses yang akan diimplementasikan menggunakan *open source tools* yaitu Pentaho Data Integration sebagai hasil akhir dari penelitian ini. Terdapat evaluasi dengan komparasi menggunakan OpenRefine yang menghasilkan hasil data *cleansing* menggunakan Pentaho Data Integration lebih unggul dalam proses pembersihan data secara menyeluruh.

Kata kunci—*data cleansing, pentaho data integration, openrefine, iterative incremental*