During sleepy industries data conversion and consolidation, there is challenge that there are different systems with same information in different format and some legacy system is still using old file format. Our target is having a central sytem where we can have a single view of critical business data from different systems.

Our first approach is to establish a glossary that represents sleepy industries key business definitions. This glossary should include all the business language that Sleepy industries uses to communicate its information with clients, vendors and all other partners. It defines what information are needed to meet business requirements. This information will be the index of data to be captured and analyzed. Ideally, business glossary should comply with standards, such as ISO/IEC 11179-4.

Using business glossary as “key”, we can consider implementing relevant key-value store technology to store information mapped to each business key in glossary. During the mapping process, we need to classify and tag sensitive data such as credit card information, client address and other private information. After the mapping process, all business keys should be in different variable groups. They can grouped based upon source(internal vs. external), business logic, purpose and others. Grouping promote efficiency and help detect patterns.

To integrate data from different data source including legacy systems, we should consider not just convert information from one format to another. We should extract meaning from legacy system and other unstructured text to provide a 360-degree view of customers, vendors, materials, assets, cost and other entities.

Data quality during conversion has high priority. To ensure information integrity and unique, we need to follow some disciplines that includes methods to measure, certify the quality and integrity of data. We need collaborate with business stakeholders to develop a quality requirement matrix to identify critical data elements, data quality concerns and business rules. Based on the requirement, we should weight keys in business glossary per their criticality. To efficiently inspect the accuracy of the data, we need build summary statistics, data ranges, etc into the data dictionary to serve as benchmarks. Automatic software can be used to perform sanity checks against summary statistics. We should implement different independent checks and regularly audit the integrity of data.