



Case Study: Data Consolidation

Industry: Public Sector

Customer:
Cabinet Level Federal Agency

Challenge:
This agency needed to consolidate its data to have better and timelier access to information for agency-wide reporting. The large amount of information, however, was a concern in that the costs and resources necessary to complete the project were immense.

Solution:
The agency and its independent consultant chose Sypherlink's Harvester to automate the discovery and mapping process and dramatically reduced the time and number of resources needed to complete the consolidation. As a result, the project was completed at a fraction of the originally estimated cost.

Cabinet Level Federal Agency Utilized Sypherlink Harvester™ to Enable Information Sharing Across Multiple Databases

As part of an enterprise architecture project, this federal agency needed to combine data from 17 databases containing 460 tables and 7,233 columns. With this volume of data, cost, resources and time were a huge concern. In order to ensure success, the agency hired a third-party consultant firm with a history of data integration success to help with the technology selection and execution of the project. The firm's task was to conduct a pilot project with the data that would serve as a prelude to a project of a much larger scale, including additional data sources from several areas across the country. The goal was to consolidate product summary information that spanned the multiple data sources.

Using Technology to Replace Manual Effort

The agency had originally estimated that it would take several months of manual mapping to complete mapping process needed to consolidate the information. It realized quickly that it needed to find a way to shorten that process. There was also some concern for the integrity of the data and maintenance costs given the known error rates related to human mapping projects.

After some evaluation, the agency decided to utilize Sypherlink's unique technology, Harvester, to automate as much of the mapping process as possible. The implementation team divided up the data mapping effort into several phases:

- Set Up – Installation and configuration
- Discovery – Identify metadata and schema for defined data sources
- Heuristic Matching - Uncover inter- and intra-database field level relationships based on patented heuristic matching technology which includes an analysis of the schema information as well as the data itself
- Relationship Filtering – User applied domain expertise to mask non-applicable relationships
- Exception Mapping – Paring down the relationships to show only what was needed to support the roll-up reporting



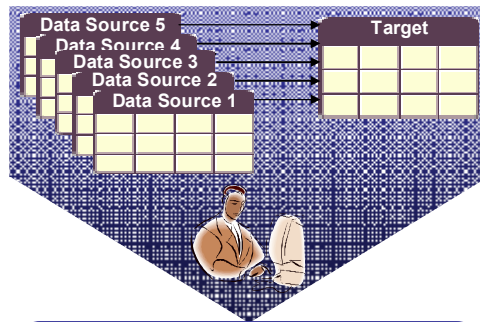
sypherlink™

Knowledge • Access • Reliability

Results:

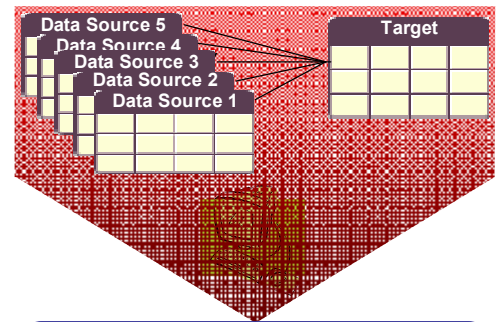
- Dramatically reduced mapping effort and cost
- Shortened project timeline by months
- Higher quality of maps
- Standardized data for use in future data-sharing initiatives

Manual Mapping Process



- ✓ Can only interrogate one data source at a time
- ✓ Maps 1.5-2 relationships per hour per person
- ✓ Extremely resource intensive
 - Resources are costly
 - Resources often change
- ✓ Highly error prone process
- ✓ Ongoing maintenance has similar time and resource constraints

Using Harvester



- ✓ Maps 60-150 relationships per hour
- ✓ Performs automated metadata discovery
- ✓ Utilizes sophisticated heuristic matching
 - Field name analysis
 - Field attribute analysis
 - Referential data analysis
 - Content analysis
- ✓ Exceptions mapped through intuitive GUI
- ✓ Interfaces to 3rd party data management tools

Sypherlink Harvester dramatically reduced manual effort, time and cost in the agency's data consolidation project.

Automation Lowers TCO

The process was completed in 52 hours, only two of which required human intervention; the remaining time was spent in unattended processing. Only one resource with the appropriate domain expertise was needed to assist in set-up and various mapping verification tasks. The agency had originally estimated that it would take several months of manual mapping to complete what Sypherlink's technology completed in two days. The agency was also prepared to staff a project team to research data sources, identify relationships and document the process to prepare the data for the target repository.

The time, resource and cost savings associated with this project are obvious from the facts provided. However, the intangible benefits are also very valuable. This technology provided a higher quality of maps than would have been possible with the normal rate of inconsistencies inherent in human intervention. There is also more flexibility with the agency's data as the completed mappings can be used for many other purposes including data translation, data migration, data federation, business intelligence, reporting, data quality, and other data-sharing initiatives.