

## Data Mapping

### DATA MAPPING

Find out more about navigating Wikipedia and finding information •

Data mapping is the process of creating data element mappings between two distinct data models. Data mapping is used as a first step for a wide variety of data integration tasks including:

1. Data transformation or data mediation between a data source and a destination.
2. Identification of data relationships as part of data lineage analysis.
3. Discovery of hidden sensitive data such as the last four digits social security number hidden in another user id as part of a data masking or de-identification project.
4. Consolidation of multiple databases into a single data base and identifying redundant columns of data for consolidation or elimination.

For example, a company that would like to transmit and receive purchases and invoices with other companies might use data mapping to create data maps from a company's data to standardized ANSI ASC X12 messages for items such as purchase orders and invoices.

X12 standards are generic Electronic Data Interchange (EDI) standards designed to allow a company to exchange data with any other company, regardless of industry. The standards are maintained by the Accredited Standards Committee X12 (ASC X12), with the American National Standards Institute (ANSI) accredited to set standards for EDI. The X12 standards are often called ANSI ASC X12 standards.

In the future, tools based on semantic web languages such as RDF, the Web Ontology Language (OWL) and standardized metadata registry will make data mapping a more automatic process. This process will be accelerated if each application performed metadata publishing. Full automated data mapping is a very difficult problem (see Semantic translation).

### Hand-coded, graphical manual

Data mappings can be done in a variety of ways using procedural code, creating XSLT transforms or by using graphical mapping tools that automatically generate executable transformation programs. These are graphical tools that allow a user to "draw" lines from fields in one set of data to fields in another. Some graphical data mapping tools allow users to "Auto-connect" a source and a destination. This feature is dependent on the source and destination data element name being the same. Transformation programs are automatically created in SQL, XSLT, Java programming language or C+++. These kinds of graphical tools are found in most ETL Tools (Extract, Transform, and Load Tools) as the primary means of entering data maps to support data movement.

### Data-driven mapping

This is the newest approach in data mapping and involves simultaneously evaluating actual data values in two data sources using heuristics and statistics to automatically discover complex mappings between two data sets. This approach is used to find transformations between two data sets and will discover substrings, concatenations, arithmetic, case statements as well as other kinds of transformation logic. This approach also discovers data exceptions that do not follow the discovered transformation logic.

### Semantic mapping

Semantic mapping is similar to the auto-connect feature of data mappers with the exception that a metadata registry can be consulted to look up data element synonyms. For example, if the source system lists FirstName but the destination lists PersonGivenName, the mappings will still be made if these data elements are listed as synonyms in the metadata

## Data Mapping

registry. Semantic mapping is only able to discover exact matches between columns of data and will not discover any transformation logic or exceptions between columns.