



**Data wrangling for
effective metadata
cleansing and mapping**



Organizations utilize many systems to store, access and retrieve information. To do this, end users search these systems by utilizing system attributes which are populated with critical metadata (e.g., document origin, or classification). Metadata describes the who, what, where and how of the stored information.

System attribute requirements differ, with the accuracy and reliance of the populated attribute being of utmost importance to allow for identification of information to perform activities.

Organizational challenge

Unfortunately, system attribution is not always mapped correctly to the original source information with accurate metadata, which can lead to utilizing and sharing incorrect information.

Organizations also face the challenge of migrating information from and to systems, or implementation of new numbering schemas, which will require substantial mapping and cleansing activities to reflect source information correctly. When not managed and implemented correctly, this can contribute toward inaccurately representing information.

When system attribution is missing or inaccurately mapped to system attribution, this contributes toward:



Duplicate information.



Inability to identify, utilize and share information.



Additional man-hours spent trying to locate information.



Utilization of personal drives because of untrustworthy and inaccurate systems.



Enhanced risk of working to historical information, thus contributing toward personnel near-miss and incidents.

There is also the additional challenge of migrating information from and to systems, where organisations will need to manage:



Identifying tool to migrate and load information en masse.



Mapping metadata to new system attribution requirements.



Maintaining legacy metadata/attribution within new systems to enhance search capabilities and maintain origin information.



Monitor migration implementing quality checks which can impact personnel workloads.

Data wrangling services

Data cleansing and mapping technologies need to accurately detect, correct, eliminate and transform metadata to align with source information and system attribution requirements.

Data wrangling incorporates adaptable technologies that enable the cleansing and mapping of metadata extracted from systems and documents, allowing for alignment to system attribution and global taxonomies. This includes detection and alignment inconsistencies which may have been originally caused by user entry errors or differing definitions of similar entities.

With data wrangling services, organizations are able to consider mapping, retaining and managing metadata such as:

- Associated document relationships, like appendices
- Document to MOC relationships
- Document to Tag relationships
- Document to Equipment relationships
- Document to Process Unit relationships
- Document to Purchase Order relationships
- Tag – Equipment relationships
- Purchase order to Tag relationships

After cleansing and mapping activities, organizations can ensure that data is consistently applied and compatible with system requirements. The data will also be transformed into a system compatible load sheet.

Data wrangling approach

Data wrangling services coupled with vision analysis, deep learning, machine learning technologies and domain SME engineering IMDC knowledge enable 'en masse' cleansing and mapping of information and its associated data.

Data wrangling processes will be adaptive and configurable to organizational requirements. This enables seamless data transformations and standardization activities, such as:

- Analysis of document and system metadata extracts
- Identifying gaps and missing or erroneous data
- Review of target system mandatory metadata fields and expected data for each field
- Alignment of organizational numbering specifications and/or procedures with system attribution
- Implementation and management of transformation scripts and gap analysis on results of the aligned and unaligned data
- Identification of value-add opportunities with available attributes for enhancement of data

capabilities (e.g. System Numbers, Area Codes, MOC Numbers, Dates, etc.)

- Aligning identified value-add opportunities with organizational numbering specifications and/or procedures
- Implementation of specialized system provision mapping scripts to clearly define attribute formats, such as date formats and removal of white space or illegal characters, etc.
- Aligning data with target system list of values and transform data into a system compatible load sheet

The integrity of data would not be compromised during cleansing and mapping activities, which would also be closely monitored with key metrics.

The result

Mapping, cleansing, and transformation of data is a pre-requisite for accurate identification of critical data. Data wrangling ensures that organizations work and share true source data while reducing the likelihood of risk to personnel, incidents, or downtime and respective cost impacts.

About the author

Janine Murray

Consulting Practice of Energy, Natural Resources, Utilities and Engineering & Construction

Janine Murray is an IM Consultant with over 15 years of experience in the O&G industry. She has extensive FE/Operations and Major Capital Project (MCP) Information Management experience. She also possesses deep experience with IM brownfield modifications, greenfield enhancements, MCP joint ventures, Closeout, and MCP handover to Operations. Additionally, she is experienced with document cleansing and data extraction techniques for digitizing O&G legacy assets.

She can be reached at janine.murray@wipro.com



Wipro Limited

Doddakannelli, Sarjapur Road,
Bangalore-560 035, India

Tel: +91 (80) 2844 0011

Fax: +91 (80) 2844 0256

wipro.com

Wipro Limited (NYSE: WIT, BSE: 507685, NSE: WIPRO) is a leading global information technology, consulting and business process services company. We harness the power of cognitive computing, hyper-automation, robotics, cloud, analytics and emerging technologies to help our clients adapt to the digital world and make them successful. A company recognized globally for its comprehensive portfolio of services, strong commitment to sustainability and good corporate citizenship, we have over 175,000 dedicated employees serving clients across six continents. Together, we discover ideas and connect the dots to build a better and a bold new future.

For more information,
please write to us at
info@wipro.com

