

A woman in a black business suit and glasses stands next to a man in a white shirt and tie. They are in a server room, looking at a laptop held by the man. The room is filled with server racks and cables.

Data wrangling for effective metadata extraction

Organizations are realizing that there is valuable information hidden within their legacy files (both paper and electronic) which can now be efficiently identified due to improved methods for extracting metadata. Metadata has several forms and can be located within many areas. For example, in the oil and gas industry, metadata contains information pertaining to fields, basins, geological provinces, etc. In addition to the company, third parties may use or need the metadata - contracting companies, vendors, fabricators and partners.

The importance of accurate and accessible metadata should not be undervalued. Accessible metadata attribution provides end users with confidence that the origin, history, and integrity of each attribute comes from a reliable source.

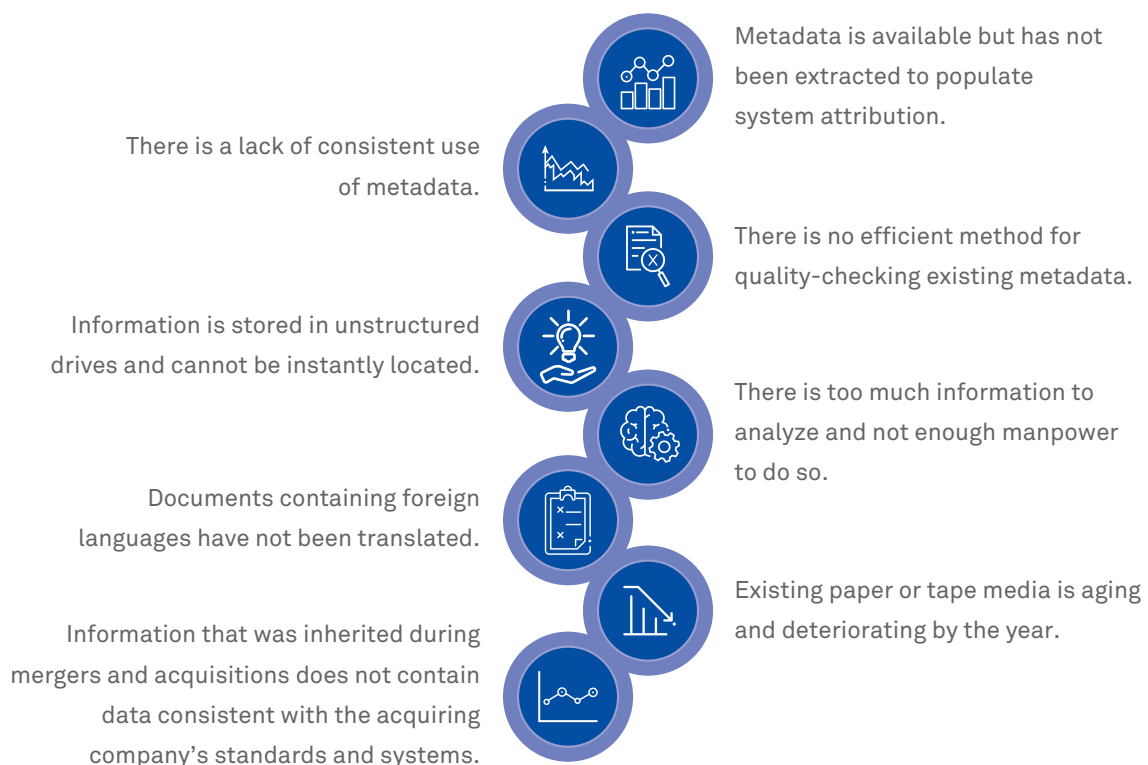
Document metadata is also utilized to populate system profiles for easy end-user access and less end-user copies.



Organizational challenge

Many organizations struggle to enforce the information management standards, processes or procedural guidance for metadata-often due to cost or resource constraints. However, with the right course of action, companies can unlock the value in previously inaccessible metadata and documents.

There are many metadata challenges that organizations are currently facing today, such as:

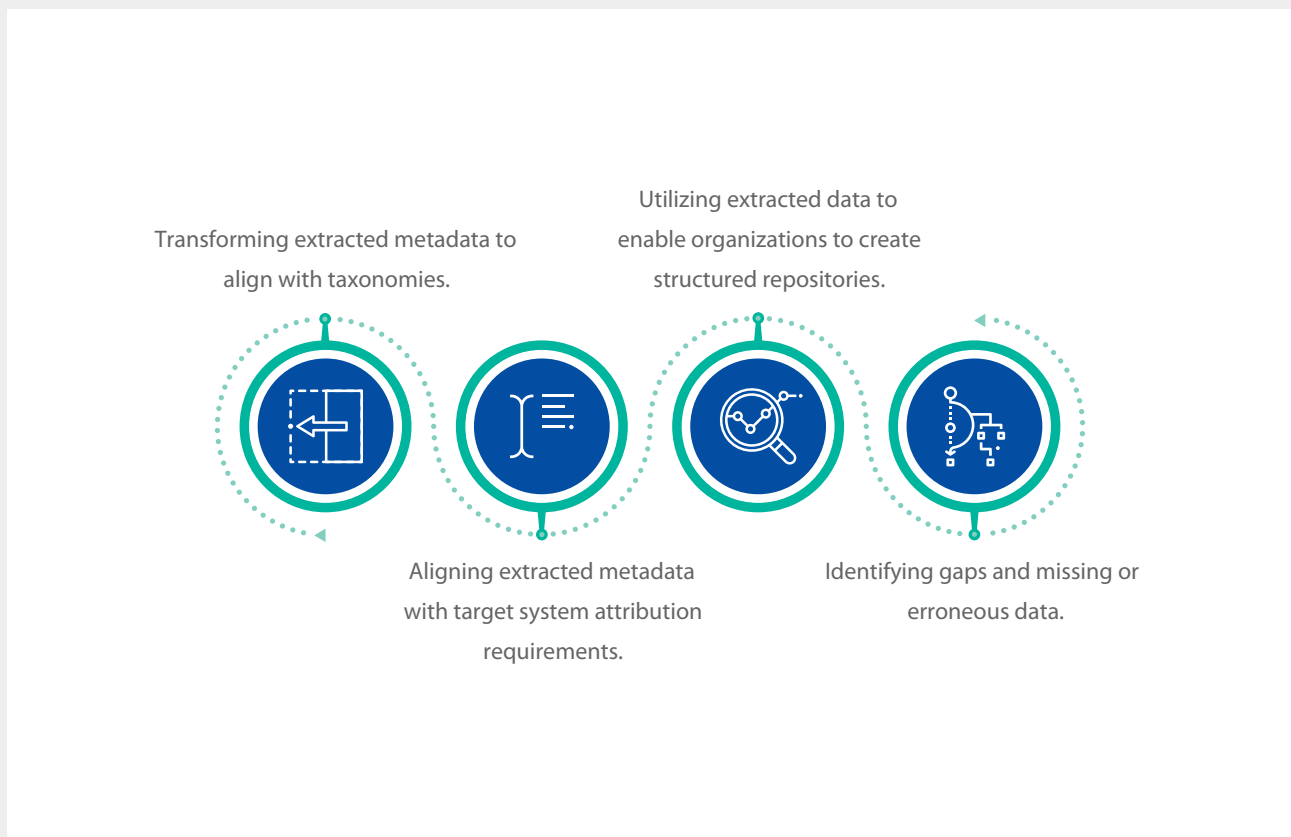




Data wrangling services

Data wrangling services support an organization's drive to digital by extracting accurate and complete metadata from previously inaccessible repositories of information. Furthermore, metadata associations and relationships between document revisions, attachments or mark-ups/comment sheets must be maintained.

Data wrangling solutions demonstrate an intimate understanding of the inherent challenges of digitizing data, as well as a scalable, structured and phased approach to metadata extraction. Metadata extraction utilizes targeted technologies to accomplish data extraction, such as computer vision, machine learning, and text analysis. Additionally, SMEs' hands-on experience supports organizations drive to digital by maximizing the value of extracted metadata, such as:



Data wrangling approach

Metadata extraction and cleansing activities can be customized by defined organizational rules, such as numbering and coding procedures, target system attribution requirements, etc. Data wrangling services would encompass several metadata extracting techniques, such as:

- **Exact Phrase:** Intended for items like a list of Originators, Country, Basin, Field

- **Key word trigger:** Intended for items such as Title
- **Form based:** Used for repetitive forms to extract targeted information
- **Vision analysis based:** Used to dynamically find an area on the document and extract targeted information
- **Free form extractions:** To extract targeted data

By leveraging domain SME engineering IM knowledge, extracted metadata can be processed, mapped and aligned to organizational requirements, such as:

- Identifying gaps and missing or erroneous data
- Reviewing mandatory system metadata field requirements and expected data for each field
- Aligning metadata to numbering Specifications and/or Procedures
- Reworking metadata transformation to include newly aligned codes and ensure no gaps are present
- Identifying value-add opportunities with available attributes for enhancement of data capabilities (e.g. System Numbers, Area Codes, MOC Numbers, etc.)
- Aligning identified value-add opportunities with global taxonomies



The result

The ability to extract rich and searchable data from previously inaccessible repositories can transform the way a company operates. Implementing data wrangling services can lead to:

- Increased data integrity
- Enhanced safety
- Streamlined processes
- Increased data traceability
- Improved decision quality
- Accelerated decision-making
- Increased operational use of information that was previously unreachable

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 has 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

