

Data wrangling
for effective
file quality correction
& classification



Organizations create, receive and share significant quantities of files and data during any given project, operation or exploratory activity. Unfortunately, the quality of each file received can vary, with quality issues usually arising when the file originator is no longer under contractual obligation to the organization.

Re-creation of information may seem like the immediate solution but this usually is not a long-term fix and can add to files retention issues, duplication, and cost. All resulting in potential future issues with the ability to create significant impacts.

Organizational challenge

We are in the middle of a data explosion where the world is estimated to have added 90% of all available data in the last 2 years. We create 2.5 quintillion bytes of data each day at our current pace – resulting in real organizational challenges:

- Moving legacy data (e.g., paper, old files) into current technologies
- Excessive man-hours spent trying to locate information
 - Cost implications of unproductive hours being spent interrogating systems and unstructured areas
- Data completeness
- Data usability
- Duplicate information

Data wrangling services

Dependable and trustworthy quality correction software and processes can be tailored to organizational requirements to safeguard the integrity of their files and the accuracy of their data. Data wrangling services partnered with domain SME experience allows for the implementation of effective file enhancement processes, thus ensuring high-quality files are searchable; allowing systems to fully digest their content.

Data wrangling services also provide the ability to classify information appropriately; aiding findability and completeness checks.



Data wrangling approach

Quality correction and classification utilizes a combination of techniques, including vision analysis, machine learning, deep learning, and

deep domain expertise to truly accommodate the complexity of today's information. These techniques enable key results:

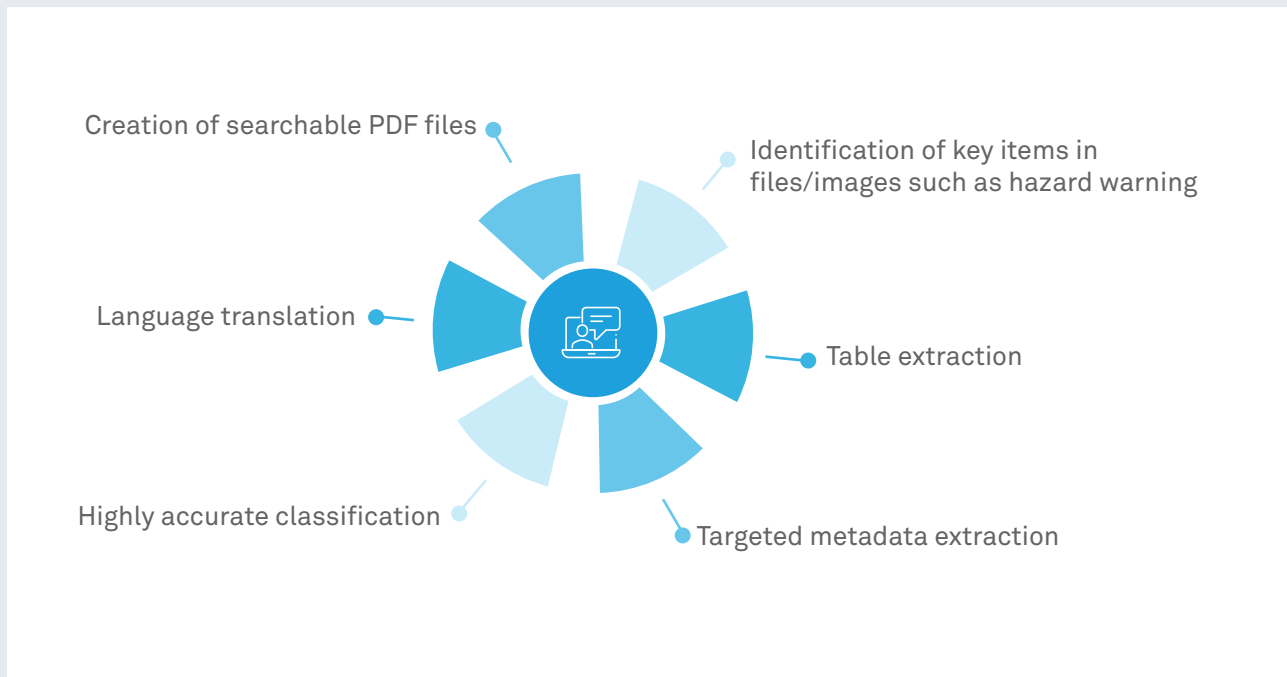


Figure 1: Data wrangling approach

The result

Data wrangling addresses quality and classification issues, which maximizes information identification. This maintains the

integrity of the files, ensuring their accuracy, consistency and reliability.

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

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

