

Data wrangling for normalizing files



Global organizations maintain an unquantifiable amount of inaccessible files which contain pertinent inaccessible metadata. Inaccessible data repositories are mostly ‘out of sight – out of mind’, with organizations not knowing the exact volume or value contained within these repositories.

There is significant value that can be retained in these repositories without redrafting documents, which can prove to be an expensive affair. There is no value in retaining data that is not utilized. One aspect of releasing the value is to normalize files. This process entails bringing them up to current technology standards.

Organizational challenge

Unfortunately, poor quality of information can only worsen over time. Organizations, today, are facing the very real challenge of identifying files contained within structured and unstructured environments. These may be inaccessible and illegible for several reasons, including but not limited to:



Variation in preparation and scanning induces notable systematic differences



Obvious “streaking” and a variation in background intensity



Stained scanned material
e.g. coffee stains



Original color documents that have been scanned and the color overrides the text e.g. red well logs



Low-quality images prohibit end users from searching for critical data



Non-searchable files



Other company-specific issues

Data wrangling services

Advances in technology today are helping organizations unlock hidden value. Businesses are finding that data wrangling techniques, coupled with artificial intelligence, leads journey to unlocking and harnessing trapped metadata.

Data wrangling normalization methods, approaches and techniques are designed to specifically prevent destructive impact i.e. prevent the loss of data. For example, where the original data is moderately good, normalization techniques gently fine-tune data to ensure that raw data is not drastically affected. Further, where original data is of poorer quality, (e.g. extremely dark background) advanced techniques separate the elements for individual treatment and then recombine, thereby resulting in an optimal update.

Data wrangling approach

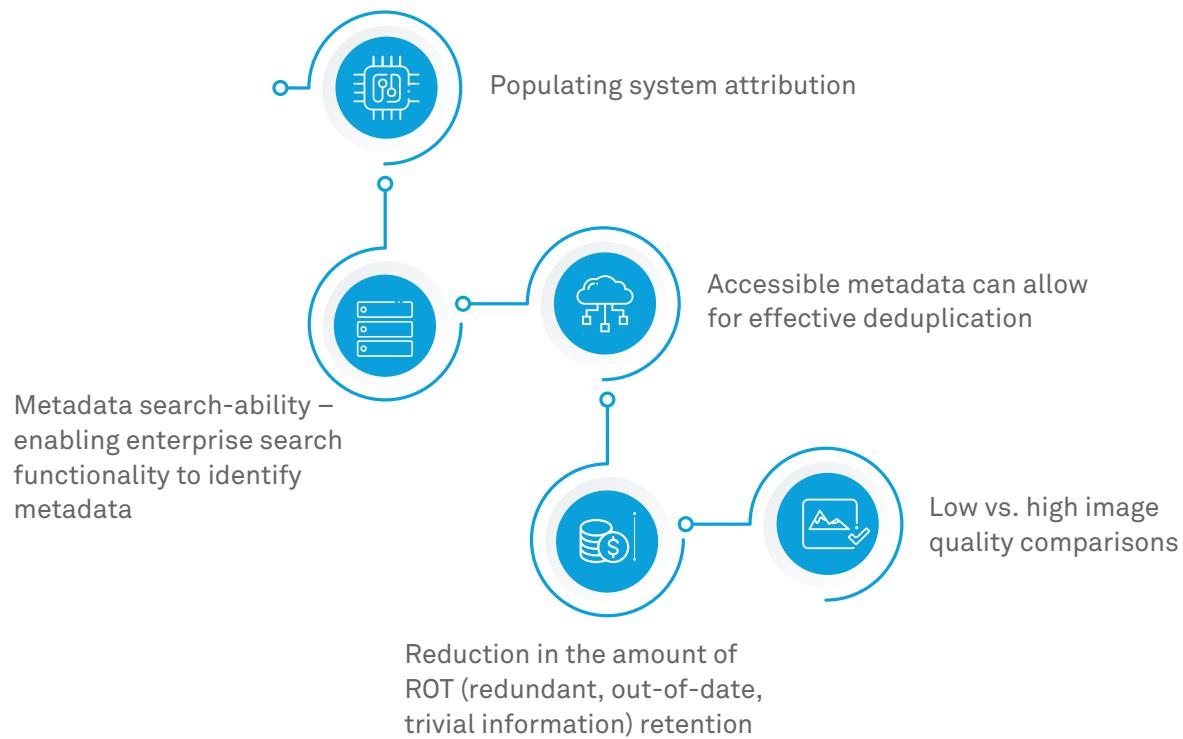
Organizations require a solution that is tailored to their needs without compromising the integrity of files and associated metadata. Specialized, normalising, pre-processing data wrangling techniques addresses the following:

- Skewed images - to correct squint scans
- General background noise - such as eliminating grainy dots
- Blurry or unclear images - sharpens text and lines on image
- Poor color - color adjustment
- Poor brightness or contrast - corrects brightness and contrast
- Non-searchable PDFs - converted to searchable

The result

Normalising data releases value which was once illegible, enabling accurate and quicker decision-making within the organisation by maximising data's search-ability.

Data wrangling features need to be developed and based on an intimate understanding of the inherent difficulties associated with normalizing files. This is the first step to releasing hidden value, which can in turn contribute towards:



About the author

Janine Murray **Consulting Practice of Energy, Natural Resources, Utilities, Engineering and 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 extensive experience in IM brownfield modifications, greenfield enhancements, MCP joint ventures, Closeout, and MCP handover to Operations. Additionally, she is experienced in 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