ASSET INTEGRITY MANAGEMENT – HARVESTING RECORDS & DATA

RadixData's clients are operators of liquids and gas pipelines, gas plants, refineries and chemical plants, as well as the engineering services companies and enterprise software companies involved in the projects. RadixData supports projects such as MAOP for liquids and gas pipelines, Facility "Fitness for Service", Corrosion Reviews, and Risk-Based Inspection Reviews.

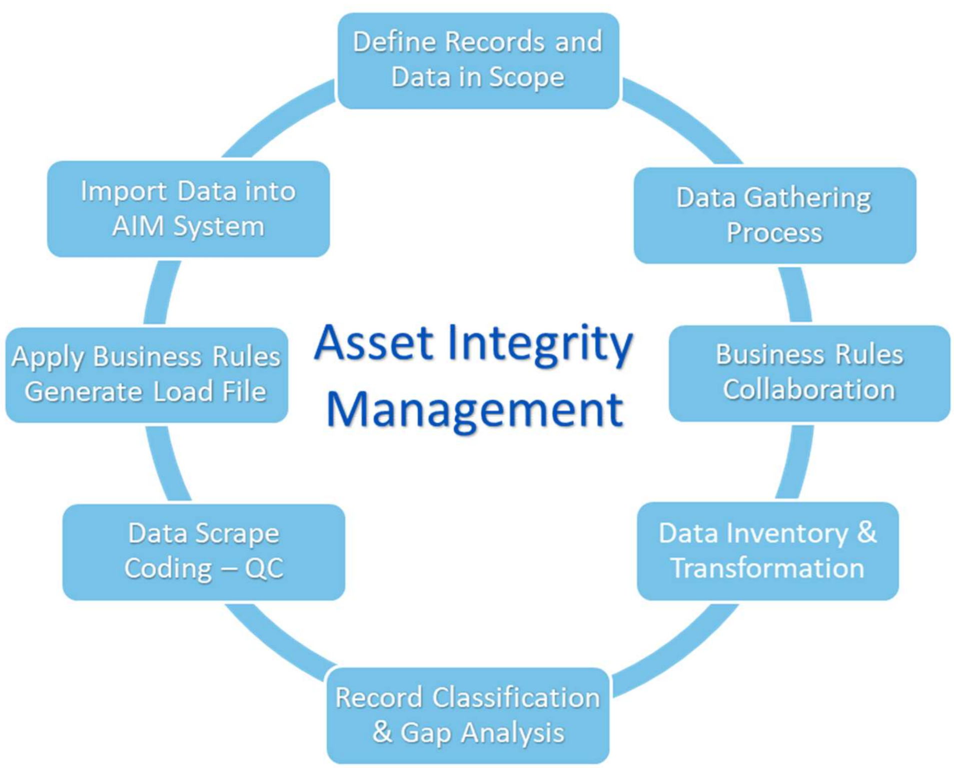
Several high-profile events (and numerous smaller events) have increased the pressure on operators to analyze correction trends and safe operating pressures. Often locating, organizing, classifying and indexing the data becomes the critical path challenge for Asset Integrity Management systems which depend on large amounts of high-quality data as an input.

Operation of these assets falls under several regulatory bodies including the Department of Transportation (DOT), Environmental protection Agency (EPA) and the Occupational Safety and Health Administration (OSHA). In accordance with the regulations, operators must have procedures and standards (such as ASME and API) for ongoing maintenance, inspection and reporting obligations associated with operation of their infrastructure with an emphasis on cost and risk management. Key to the mission is the locations and harvesting of design of materials, original construction, repair, testing and corrosion history data in order to establish a baseline of risk.

Operators of midstream and downstream facilities including pipelines, refineries and chemical plants seek to ensure the safe and efficient operation of their infrastructure and compliance with federal regulations relating to public safety. In addition, operators seek to improve maintenance planning and CAPEX planning commensurate with risk analysis. The affected infrastructure is often decades old and has been subject to numerous merger and acquisition events, which has resulted in the creation of silos of poorly understood and indexed "dark data", as well as the effects of an aging workforce who possesses the last remnants of historical understanding of the facilities.

RadixData's job is to rationalize the data to the objectives of the engagement and deliver a clean data set which can feed other applications.

OUR PROCESS



DEFINING RECORDS AND DATA IN-SCOPE

Asset Integrity Management requires data from multiple sources some of which are vintage paper records and antiquated or proprietary systems. RadixData has a process for identifying those sources, analyzing the content and developing a process for mining data from those sources. The process is completed when all data is gathered, and a complete data set exists to import into an Asset Integrity Management System. Our clients leverage our expertise in filling addressing the problems with gathering vintage records and data to fill the gaps of missing data. The process defining of records and data in-scope the gathering of the data is iterative and continues until all gaps in data are filled.

DATA INVESTIGATION AND COLLECTION

Radix Data will organize interviews with key stakeholders in the operator's organization to fully understand record descriptions, search criteria, physical and virtual locations for data and records of interest. The interview yields information held by long time employees and client subject matter experts, providing a history of the types, locations, and nomenclature associated with legacy records. This input feeds the collection process, where collection teams are fielded to defensibly query and gather both physical and electronic records. Collected records and data are processed in our facilities to populate the project database.

CONTENT ANALYTICS

Radix Data performs "Records Viability Analytics" of current metadata to assess the availability of critical records types. Although generally sparse and incomplete, the metadata hits will point to the "low hanging fruit" records which appear to be readily available. This metadata typically includes digital and physical file inventory descriptions and searchable index fields in content management systems. Performing analytics on the metadata creates a profile of the available records, facilitating decisions by the project team to intelligent processing.

DATA MINING OBJECTIVES

The data mining objectives associated with preparing a highly curated engineering data set for an AIM program typically consist of the following components:

- Establishing a data schema which meets the needs of the engineering and risk analysis
- Locating and managing high value records across silos of physical and electronic repositories in a cost-efficient manner, where the records are usually poorly indexed and not searchable
- Locating complementary data from current and legacy structured data which can be used as validation and/or lookup tables for correlating data
- Identifying gaps in record coverage and establishing a remediation plan to identify or generate missing records
- Applying business rules (ex: convert fractions to decimals), low-level engineering tasks (ex: lookup applicable ASME codes) and data normalization.
- Producing a database containing the data elements collected from the available documents and databases
- Populating software applications including risk-based inspection (RBI), plant management systems (PMS) and geographical information systems (GIS),
- Performing trending and anomaly analysis of such as vessel and pipeline inspection points and corrosion
- Achieving ROI on investments in software and services to perform risk assessment

THE DATA GATHERING PROCESSES

Our observations of the difficulties associated with current data gathering processes include the following:

- Poor knowledge of available records needed to complete the engineering and risk assessment
- Silos of on-site physical records and corporate repositories of images which are poorly organized and indexed
- Poor quality record images (PDF and TIFF) which are not searchable
- Large volumes of components (vessels, piping, valves, meter stations, pipeline) which require engineering review
- Engineering resources are used to perform manual tasks to locate records and perform data entry
- Poorly defined quality control processes to ensure data integrity
- Antiquated or Obsolete Systems containing valuable information
- Proprietary Systems holding data hostage
- Data is spread across many databases, excel spreadsheets or other systems

BUSINESS RULES COLLABORATION

The development of Business Rules is the foundation for creating complete and defensible data for ingestion into any asset management system. Business Rules clearly define the process for collecting data, validating said data and standardizing the data making the process reliable. The process is repeated for each attribute and therefor traceable. Business Rules drive the records collection process as data gaps are filled. Collaboration to develop the business rules begins at the onset of the engagement and continues throughout the process.

DATA INVENTORY AND TRANSFORMATION

Source Data from share drives containing PDFs, excel spreadsheets, databases and/or digital images are processed through our Data Inventory and Transformation Process where has tags are generated for chaing of custody, zip file content is extracted and flattened for inventory of each file, removal of blacklisted file types, de-duplication are performed. The inventory allows for tracking the files and their pages throughout the entire processing cycle.

RECORDS CLASSIFICATION

RadixData performs discreet records classification after the DIT process where we assign each document to a classification based on title and/or the structure of the content on the document. Where a title is available and can be matched, the document is identified as belonging to a particular classification. While your organization may contain many unstructured sources, it’s unlikely that all the information is valuable to your organization. By using our services and the power of computer processing, the attribute extraction platform can sort through and find **only** the information you need. Think of it as an automatic filter for any unstructured data that you are managing.

GAP ANALYSIS

Once classification is complete, Radix performs a Gap Analysis to determine the presence or absence of key documents. This is a key step in the process. Often, we find that organizations feel their data is complete when in fact they are missing data. The gap analysis leverages harvested data from the documents and the naming methodology of electronic records to confirm if a document is labeled correctly. The process occurs in tandem with the data scraping and quality check processes and is repeated until all desired records and data points are present. In some instances, the primary record may not be present, but the gap analysis is satisfied according to the business rules by secondary and tertiary documents. The process is iterative and is determined as complete by our client.

ATTRIBUTE EXTRACTION

RadixData's attribute extraction platform can find data like a human would: quickly, accurately, and within unstructured documents. A typical challenge of unstructured data is that you cannot predict where the required data elements will be. Like a human, the attribute extraction software looks through documents and data sources of all kinds to find and capture the information that is important to your organization. Automating your organization's data capture process is fast, accurate, and is not limited to specific forms, nor does it require the maintenance of templates or barcodes. Documents of all kinds can be reviewed in this process. Our intelligent algorithm is made up of rules written with the **experience of processing millions of documents** and unstructured data, similar to what your struggles are today. The platform can easily be enhanced, via a customer specific layer, taking advantage of all the existing knowledge and configuring new rules specific to your organization's needs.

PRE-VALIDATE

A powerful feature of the extraction platform is the ability to pre-validate information found in your unstructured data. Our platform can match existing information within your database. The system also has the ability to pre-validate against standard logic or criteria defined by your business rules. Our automated data extraction software is able to help your organization achieve the highest level of accuracy and efficiency by eliminating steps within current workflows.

VERIFICATION

Where verification is required, the verification user interface has been developed to reduce errors, reduce eye strain, and point out specific areas where verifiers should focus attention or where the platform is confident the data is correct. The ease of use and productivity tools are designed to be intuitive which makes shorter training requirements and easier on-boarding of new employees.

DATA STANDARDIZATION/NORMALIZATION

Once data has been extracted from unstructured sources, RadixData then standardizes and normalizes it, then saves the captured information. A major strength of the platform is the ability to easily format its output to meet the specifications of any downstream system. If unstructured data is arriving from a different department or external organizations, it often does not conform to formats or naming conventions your organization uses. As a standard part of the data extraction process, RadixData's platform can automatically map extracted data to conform to your organization's naming conventions or simply reformat data to meet specified date, number or other format requirements.

CUSTOMIZABLE WORKFLOWS

Another common burden of an organization's data is it typically must be manually handled by a plethora of employees throughout an organization. Not until all necessary employees have reviewed the data can the relevance, value, retention and disposition be determined. Customizable workflows allow the administration of the review in an efficient manner by developing a workflow to send any multitude of alerts to users and/or administrators in order to bring items to the appropriate level of attention right away. To keep track of processing time, reduction of error-rates, and increases in employee productivity, RadixReview has built-in reporting dashboards. If preferred, the RadixReview platform can output any data (usage and content) to a shareable format.

SPEED & SCALABILITY

We retrieve data from your documents quickly and reduce the time it takes to deliver a complete and verified data set. Our platform has a very predictable resource curve and can accurately estimate and scale up to process more documents at any given time.

PROFESSIONAL SERVICES

Each customer has a designated support services representative that will work hand-in-hand to optimize, roll-out to additional departments, upgrade, and troubleshoot any issues customers may be experiencing. Scheduled bi-weekly meetings are held to deliver quarterly dashboards and hold customer roundtables, as well as to touch base and answer any technical questions customers might need help with.