Overview

MARIA is a government transparency platform that is built to simplify how public data is processed and acknowledged in the Dominican Republic. It automatically extracts the data from different official sources, then reviews it by using AI and presents it in a clear and searchable format.

Before MARIA, procurement information was scattered across different websites, hidden in complex documents, and was very tough to track over time. Users had to search the data manually, download and interpret data, which is the quiet hectic process.

By automating data collection and analysis, MARIA now enables users to access procurement insights within minutes instead of days, improving transparency and making government spending more visible and easier to understand.

Company Background

MARIA operates at the intersection of public data, transparency, and technology. It is extremely beneficial for the researchers, journalists, and advocates who need access to authentic government data.

Government procurement systems generate data daily in bulk amounts, but the information is not properly structured. This is challenging for the extractor to extract the right information or identify the patterns. MARIA was designed to compensate for this gap by converting the raw data into the usable and structured form.

Problem Statement

1. **Fragmented Data Sources**

Procurement data was spread across multiple government platforms, making it difficult to track all tenders in one place.

2. **Unstructured and Inconsistent Documents**

Documents are available in different formats such as PDFs, scanned files, or HTML pages, but it is not present in an approved way.

3. **Limited Data Accessibility**

Users had to manually search and review documents, often missing important information due to time constraints or complexity.

4. **Lack of Company and Ownership Visibility**

Tracking companies and their ownership structures was nearly impossible due to inconsistent naming and lack of centralized records.

5. **Scalability Challenges**

Thousands of documents were published each year, making manual monitoring and analysis unrealistic.

6. **Delayed Insights**

Important procurement information was often discovered days or weeks after publication, reducing its usefulness.

Proposed Solution

MARIA introduced a centralized platform that automates the entire procurement data lifecycle, from collection to analysis and visualization.

ü **Automated Data Collection**

The platform continuously gathers tender data from government sources on a daily schedule, ensuring no announcements are missed.

ü **AI-Powered Document Analysis**

AI is used to extract key information such as bidders, contract values, and procurement details from unstructured documents.

ü **Centralized Data Management**

All tenders, companies, and ownership data are stored in a single system, making it easy to search and analyze.

ü **Search and Filtering Capabilities**

Users can quickly find relevant tenders, companies, or trends using filters and full-text search.

ü **Relationship Mapping**

The system identifies and visualizes relationships between companies and their owners, providing deeper insights into procurement networks.

ü **Interactive Dashboards and Reporting**

Users can view procurement trends, analyze spending patterns, and generate reports directly from the platform.

Implementation

1) **Automated Data Pipeline**

A structured pipeline was built to handle continuous data flow:

- Daily scraping of procurement websites
- Secure storage of documents in the cloud
- Monitoring system to detect new or updated files
- Background processing of documents for analysis

2) **AI Processing Layer**

Documents are processed through an AI system that:

- Extracts text from files
- Identifies key data points
- Structures information for easy use

3) **Data Storage and Organization**

Processed data is stored in a database where it is:

- Cleaned and standardized
- Deduplicated to avoid repeated entries
- Linked across tenders, companies, and ownership records

4) **User Interface and Dashboard**

A web-based interface allows users to:

- Search and filter procurement data
- View company profiles and tender details
- Analyze trends through visual dashboards

5) **System Reliability and Monitoring**

The platform includes background job processing, retry mechanisms, and monitoring tools to ensure stable performance and continuous data updates.

Results

Operational Improvements

Area	Before	After
Procurement visibility	Scattered across multiple sources	Fully centralized system
Data collection	Manual, limited coverage	Automated, near 100% capture
Access speed	Days or weeks	Within 24 hours
Company tracking	Not possible	Complete profiles with history
Ownership insights	Limited visibility	Clear relationship mapping
Reporting	Manual effort required	Instant report generation
Data quality	Inconsistent	Cleaned and standardized

Performance Gains

- Thousands of tenders processed and indexed
- Millions of data points searchable in seconds
- Real-time updates available shortly after publication

Additional Business Impact

- Improved transparency in government spending
- Faster access to procurement insights for researchers and journalists
- Better understanding of company participation and market dynamics
- Increased accountability through accessible and traceable data

Analysis

What Worked Well

- Full automation eliminated manual data collection efforts
- AI made it possible to process complex documents at scale
- Centralized data improved usability and decision-making
- Real-time updates ensured information stayed current

Key Challenges

- Handling inconsistent document formats and missing data
- Matching company names and avoiding duplicates
- Scaling the system to handle continuous data flow
- Ensuring high accuracy in AI-based extraction
- Building reliable ownership relationships across entities

Key Insights

1. Automation is critical when dealing with large-scale public data
2. AI is most effective when combined with structured data systems
3. Clean and organized data significantly improves usability
4. Transparency increases when information is easy to access and interpret

Conclusion

MARIA successfully transformed how procurement data is accessed and understood. By automating data collection and using AI to structure complex information, the platform turned scattered documents into a clear, searchable system.

This not only improved efficiency but also strengthened transparency, giving users reliable access to procurement data and helping build greater trust in public systems.