
SOFTWARE REQUIREMENTS SPECIFICATION (SRS)

Provider Data Validation & Directory Management System

VERSION: 1.0

Author: Semantic_Coders

Table of Contents

- 1. Introduction**
 - 1.1 Purpose of this Document
 - 1.2 Scope of the System
 - 1.3 Definitions, Acronyms, and Abbreviations
 - 1.4 References
- 2. Overall Description**
 - 2.1 Product Perspective
 - 2.2 Product Features
 - 2.3 User Classes and Characteristics
 - 2.4 Operating Environment
 - 2.5 Design and Implementation Constraints
 - 2.6 Assumptions
- 3. System Features (Functional Requirements)**
 - 3.1 Data Validation Agent**
 - 3.1.1 Description
 - 3.1.2 Functional Requirements
 - 3.2 Information Enrichment Agent**
 - 3.2.1 Description
 - 3.2.2 Functional Requirements
 - 3.3 Quality Assurance Agent**
 - 3.3.1 Description
 - 3.3.2 Functional Requirements
 - 3.4 Directory Management Agent**
 - 3.4.1 Description
 - 3.4.2 Functional Requirements
 - 3.5 Workflow Automation**
 - 3.5.1 Description
 - 3.5.2 Functional Requirements
- 4. Interface Requirements**
 - 4.1 User Interface Requirements
 - 4.2 Hardware Interface
 - 4.3 Software Interface
- 5. Performance Requirements**
- 6. Design Constraints**
- 7. Non-Functional Requirements**
 - 7.1 Security
 - 7.2 Reliability
 - 7.3 Maintainability
 - 7.4 Usability
 - 7.5 Scalability
- 8. Appendices**
 - 8.1 Sample Provider Data Elements
 - 8.2 Data Sources
- 9. Uses of this SRS Document**
- 10. Conclusion**

1. Introduction

1.1 Purpose of this Document

This Software Requirements Specification (SRS) document defines the functional, non-functional, and technical requirements of an AI - driven Provider Data Validation and Directory Management System.

The document serves as the foundation for system design, development, deployment, testing, and maintenance. It ensures all stakeholders share a common understanding of the capabilities, limitations, and expected performance of the solution.

1.2 Scope of the System

The Provider Data Validation & Directory Management System automates the acquisition, validation, enrichment, and management of healthcare provider information across multiple sources. The system ensures accurate provider directories, compliance with regulatory requirements, and reduction of operational overhead.

The system includes:

- Automated extraction from structured/unstructured datasets (CSV, JSON, PDFs).
- Web scraping and API-based cross-validation using public resources.
- Information enrichment from online professional sources.
- Discrepancy detection and confidence scoring.
- Centralized directory management with versioning.
- Dashboards, analytics, alerts, and reporting systems.
- Workflow automation for daily, weekly, and onboarding validations.

1.3 Definitions, Acronyms, and Abbreviations

- **AI** - Artificial Intelligence
- **API** - Application Programming Interface
- **NPI** - National Provider Identifier
- **OCR** - Optical Character Recognition
- **PDM** - Provider Directory Management
- **UI** - User Interface

1.4 References

- CMS NPI Registry API
- Google Maps API
- State Medical Board Public Lookup Websites
- Public Healthcare Provider Databases

2. Overall Description

2.1 Product Perspective

The system functions as a standalone enterprise-grade application integrating with:

- Public APIs (NPI Registry, Google Business).
- Web scraping modules.
- OCR services for document extraction.
- Internal relational databases.
- A role-based web dashboard.

It seamlessly fits into existing healthcare payer ecosystems by enabling automated data validation processes.

2.2 Product Features

- Automated web scraping and API-based cross-validation
- OCR-driven extraction from scanned provider documents
- License, certification, and credential verification
- Confidence scoring and discrepancy flagging
- Auto-generation of updated provider directories
- Review queue for manual intervention
- Analytics dashboards with validation metrics

2.3 User Classes and Characteristics

User Type	Description	Technical Skill
Data Operations Analyst	Reviews discrepancies, validates flagged providers	Medium
Credentialing Team	Oversees provider onboarding and compliance	High
System Administrator	Manages system operations, logs, and configurations	High
Business Stakeholders	Review directory accuracy, compliance reports	Low

2.4 Operating Environment

- Backend: Python
- REST APIs: Flask / FastAPI
- Database: PostgreSQL / SQLite
- Client: Web browser
- Deployment: On-premises or cloud (AWS/Azure/GCP)

2.5 Design and Implementation Constraints

- Must follow data privacy and compliance requirements
- Relies on publicly accessible data sources
- Web scraping subject to regulatory and ethical constraints
- API rate limits from third-party providers
- Only provider-related information captured (no patient data)

2.6 Assumptions

- Input datasets are synthetically generated or provider-authorized
- Public data sources are available and updated
- OCR extraction has reasonable clarity for parsing

3. System Features (Functional Requirements)

Description:

The Data Validation Agent is responsible for verifying provider demographic and contact information by cross-referencing multiple trusted external sources. It ensures that each provider's data is accurate, current, and reliable before being updated in the directory.

Functional Requirements:

1. Extract provider data elements including names, phone numbers, addresses, specialties, and practice details from input datasets.
2. Scrape provider practice websites to obtain updated contact information and service details.
3. Validate phone numbers, addresses, and location accuracy using Google Maps or similar public APIs.
4. Cross-check provider information against the NPI Registry through API calls to confirm identity, credentials, and practice locations.
5. Assign confidence scores (0–100%) to each data element based on source reliability and cross-validation results.
6. Flag mismatches, outdated entries, and inconsistencies for further review by the Quality Assurance Agent or manual validation teams.

3.2 Information Enrichment Agent

Description:

The Information Enrichment Agent enhances provider records by sourcing, extracting, and updating missing or incomplete information from trusted public and professional sources. It ensures each provider profile is comprehensive, standardized, and enriched for improved decision-making.

Functional Requirements:

1. Fetch provider credentials, certifications, professional affiliations, and licensing details from public directories.
2. Add additional information such as education history, areas of specialization, and associated hospitals or healthcare facilities.
3. Identify geographic gaps or underserved regions by analysing provider distribution and availability.
4. Generate standardized and enriched provider profiles with consistent formatting and structured data fields.

3.3 Quality Assurance Agent

Description:

The Quality Assurance Agent ensures the accuracy, reliability, and consistency of the overall provider data validation pipeline. It evaluates validated data, identifies discrepancies, and maintains quality benchmarks across all processed provider records.

Functional Requirements:

1. Identify mismatches and inconsistencies across multiple validation sources (e.g., NPI Registry, provider websites, license databases).
2. Detect outdated, inaccurate, or suspicious provider records requiring manual intervention.
3. Generate detailed quality assurance (QA) reports summarizing validation results, discrepancies, and confidence scores.
4. Prioritize high-risk provider records based on data confidence levels, impact severity, and validation outcomes.
5. Maintain historical logs of validation metrics, data quality scores, and discrepancy patterns for continuous improvement.

3.4 Directory Management Agent

Description:

The Directory Management Agent oversees the complete lifecycle of provider directories, ensuring accurate updates, efficient communication, and clear visibility into directory status and analytics.

Functional Requirements:

1. Maintain a centralized provider database with version control to track updates and historical changes.
2. Generate provider directory outputs in multiple formats, including web-based listings and downloadable PDF versions.
3. Send automated notifications to analysts for providers flagged during validation or requiring manual review.
4. Provide interactive dashboards displaying summaries, data quality metrics, validation trends, and actionable insights.
5. Maintain prioritized task queues for manual validation workflows, including supporting documentation and status tracking.

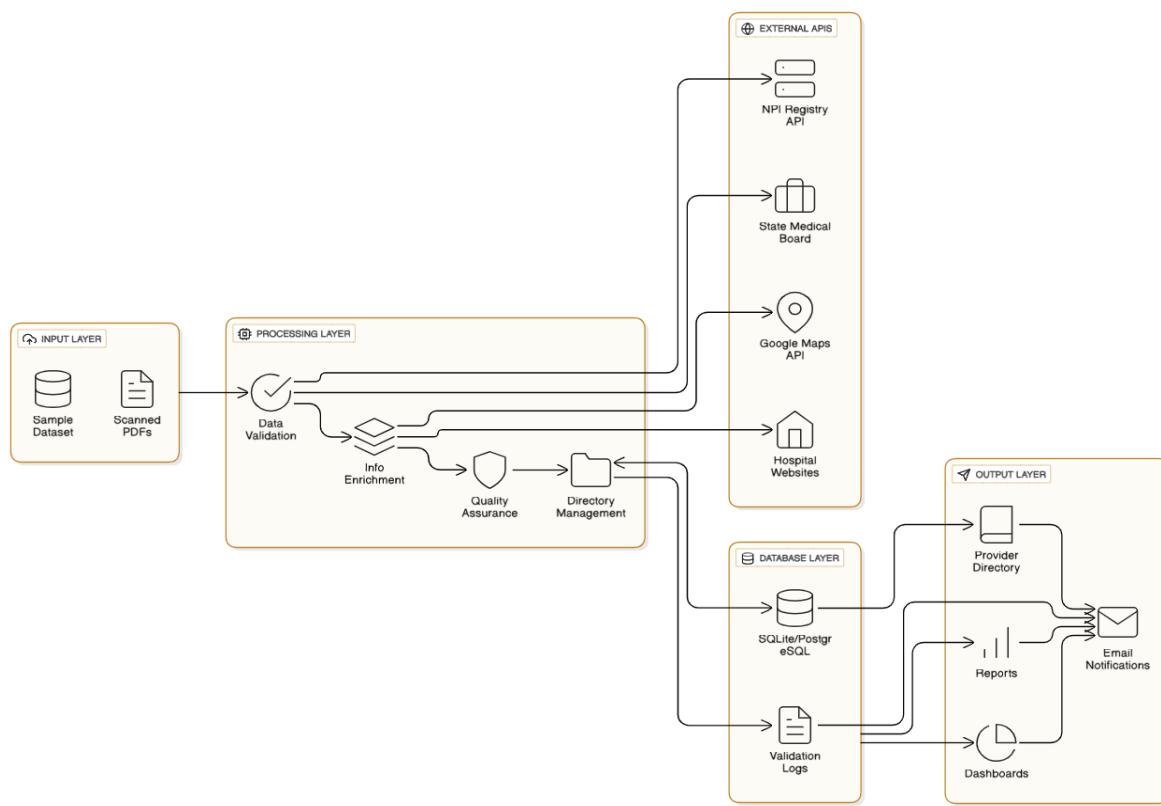
3.5 Workflow Automation

Description:

The Workflow Automation module manages and orchestrates scheduled validation tasks, ensuring continuous, timely, and rule-based execution of data validation and directory management processes.

Functional Requirements:

- Execute daily automated validation cycles for provider data accuracy and updates.
- Perform weekly provider directory health scans to assess completeness, correctness, and data quality trends.
- Conduct automated validation workflows for new provider onboarding, including credential checks and profile enrichment.
- Generate comprehensive reports for each validation cycle, including status summaries, detected issues, resolved items, and recommended actions.



4. Interface Requirements

4.1 User Interface

Dashboard components include:

- Provider List with status indicators
- Confidence score visualizations
- Discrepancy summaries

- Manual review queue
- Directory export options
- Report download section
- Dashboard analytics and graphs

4.2 Hardware Interface

- Standard PC/Laptop
- Internet connection

4.3 Software Interface

- NPI Registry API
- Google Maps API
- Web scraping utilities
- OCR/Vision model for PDF extraction
- PostgreSQL/SQLite database

5. Performance Requirements

1. Validation of 100 providers within 5 minutes (standard mode).
2. OCR extraction accuracy $\geq 85\%$.
3. Confidence scoring accuracy $\geq 95\%$ reliability.
4. Scalable for 500+ validations per hour.
5. Dashboard load time ≤ 3 seconds.

6. Design Constraints

- Must comply with healthcare data handling regulations.
- Avoid storage of any patient information (PHI).
- Must use open-source or publicly available tools.
- Scraping speed influenced by anti-bot protections.

7. Non-Functional Requirements

7.1 Security

- Role-based access control
- Secure HTTPS communication
- Automatic redaction of sensitive data

- Audit logs for all critical actions

7.2 Reliability

- Retry logic for API failures
- Robust exception handling
- Continuous logging and monitoring

7.3 Maintainability

- Modular agent-based architecture
- Configurable workflows
- Easy integration of new validation data sources

7.4 Usability

- Clean and intuitive dashboard
- Color-coded statuses
- Accessible reports and export options

7.5 Scalability

- Vertical and horizontal system scalability
- Cloud-ready architecture
- Support for thousands of provider records

8. Appendices

8.1 Sample Provider Data Elements

- Name
- Address
- Phone number
- Specialization
- License information
- Credentials
- Practice locations

8.2 Data Sources

- NPI Registry
- Google Maps
- State Medical Boards
- Hospital Directories

9. Uses of this SRS Document

This SRS will be used by:

- Enterprise development teams
- Quality assurance and compliance teams
- Product managers and business stakeholders
- Integration and operations engineers

10. Conclusion

This SRS defines a complete, scalable, and enterprise-ready AI-based Provider Data Validation & Directory Management System. It establishes requirements for ensuring provider data accuracy, compliance, reduced manual workload, standardized directories, and improved operational efficiency for healthcare organizations.