

SIMGN

Sistema de Información y Monitoreo de Gas Natural

Concurso Datos al Ecosistema 2025

Ministerio de Tecnologías de la Información y las Comunicaciones (MINTIC)

Documentación Integral del Proyecto

Información Estratégica, Técnica y Funcional

Equipo de Desarrollo

Valentina Fuentes • Raúl Lozano • Angelica Grazt

2025

Tabla de Contenidos

Introducción.....	3
Objetivo General	3
Objetivos Específicos	3
Alcance	4
Cobertura Geográfica	4
Cobertura Temporal.....	4
Usuarios Finales.....	4
Información General del Proyecto	5
Datos Básicos.....	5
Contexto y Problemática Identificada	5
Situación Actual	5
El Problema Central.....	6
Impacto de la Problemática	6
La Solución: SIMGN.....	6
Objetivos de la Solución	6
Beneficios Esperados	6
Arquitectura Técnica y Características Principales	6
Componentes Principales del Flujo	7
ETL Automatizado (El Núcleo).....	7
Almacenamiento Centralizado	7
Salidas y Visualización (Outputs).....	7
Requerimientos Funcionales.....	7
Requerimientos No Funcionales	8
Requerimientos de Datos.....	8
Stack Tecnológico y Arquitectura.....	9
Impacto y Sostenibilidad	9
Descripción del Impacto	10
Categorías de Impacto	10
Indicadores de Éxito	10
Sostenibilidad a Largo Plazo	¡Error! Marcador no definido.
Plan de Implementación.....	10
Conclusiones.....	10

Introducción

El gas natural constituye un recurso energético fundamental para Colombia, especialmente considerando la transición energética que requiere el país en los próximos años. Sin embargo, la información crítica necesaria para la toma de decisiones en este sector se encuentra fragmentada entre múltiples entidades públicas, cada una gestionando datos en formatos dispares y sin un mecanismo de integración centralizado.

El presente documento describe de manera integral el Sistema de Información y Monitoreo de Gas Natural (SIMGN), una solución tecnológica que materializa la respuesta al reto del Ministerio de Minas y Energía dentro del concurso "Datos al Ecosistema 2025" del MINTIC. Esta iniciativa se fundamenta en principios de transparencia, datos abiertos y modernización de la gestión pública.

SIMGN busca transformar la manera en que se accede, analiza e interpreta la información del sector minero-energético colombiano, creando un "Punto Único de Verdad" que democratiza el acceso a datos públicos estandarizados y permite la formulación de políticas basadas en evidencia sólida.

Objetivo General

Desarrollar e implementar una plataforma tecnológica integrada que automatice la recolección, estandarización, integración y visualización de datos públicos de producción, demanda y regalías del gas natural en Colombia, actuando como punto único de consulta para analistas, entes de control y ciudadanía en general, facilitando la toma de decisiones basada en evidencia y fortaleciendo la transparencia institucional del sector.

Objetivos Específicos

1. **Unificar las fuentes de datos:** Integrar datos de producción del MME, demanda de la UPME y regalías de la ANH en una única base de datos centralizada, eliminando la necesidad de consultar múltiples portales.
2. **Automatizar procesos ETL:** Implementar un motor de Extracción, Transformación y Carga que realice de manera automática la recolección y estandarización de datos, reduciendo errores manuales y garantizando consistencia.
3. **Garantizar calidad de datos:** Establecer validaciones robustas que verifiquen integridad, completitud y consistencia de los datos antes de incorporarlos al sistema.
4. **Crear visualizaciones integradas:** Desarrollar un dashboard interactivo que permita el análisis cruzado de producción vs. demanda vs. regalías, con segmentación territorial.
5. **Facilitar el acceso público:** Proveer una interfaz web gratuita, sin requerimientos de registro, que garantice el derecho ciudadano de acceso a información pública.

6. **Exportar datos estandarizados:** Permitir descargas de datasets en formatos reutilizables (CSV, Excel) para facilitar análisis independientes y reutilización de datos.
7. **Implementar inteligencia territorial:** Utilizar mapas interactivos para visualizar indicadores por departamentos y municipios, permitiendo análisis geográficos.
8. **Documentar y transferir conocimiento:** Generar documentación técnica completa que facilite mantenimiento futuro y adopción por otras instituciones.

Alcance

¿Qué incluye SIMGN?

- Integración de tres fuentes de datos principales: producción (MME), demanda (UPME) y regalías (ANH)
- Proceso ETL automatizado en Python con validaciones de calidad
- Base de datos centralizada con histórico de datos
- Dashboard interactivo web con múltiples visualizaciones
- API REST pública para acceso programático a datos
- Sistema de exportación automática de archivos CSV/Excel
- Documentación técnica completa
- Código fuente disponible en repositorio público

¿Qué NO incluye SIMGN?

- Datos de fuentes privadas o protegidas
- Información no pública o clasificada
- Sistemas de predicción o modelos avanzados de machine learning
- Integración con sistemas privados de empresas productoras
- Análisis regulatorio o fiscalizador automático

Cobertura Geográfica

El sistema abarca la totalidad del territorio nacional colombiano, con granularidad a nivel departamental y municipal según disponibilidad de datos en las fuentes oficiales.

Cobertura Temporal

Incluye datos históricos desde la disponibilidad en las fuentes públicas, típicamente desde 2015-2020 en adelante, con actualización automática según la frecuencia de publicación de las entidades responsables.

Usuarios Finales

- Analistas del sector minero-energético
- Funcionarios públicos del MME, ANH, UPME
- Entes de control (Contraloría, Procuraduría)
- Organizaciones de la sociedad civil
- Investigadores académicos
- Ciudadanía interesada en datos abiertos

Información General del Proyecto

Este documento consolida la información estratégica, técnica y funcional del proyecto SIMGN, participante en el concurso Datos al Ecosistema 2025 del Ministerio de Tecnologías de la Información y las Comunicaciones (MINTIC).

Datos Básicos

Nombre del Proyecto	Sistema de Información y Monitoreo de Gas Natural (SIMGN)
Entidad Retadora	Ministerio de Minas y Energía (MME)
Concurso	Datos al Ecosistema 2025 - MINTIC
Categoría del Reto	Intermedio - Entidad Pública
Nombre del Reto	Automatización e integración de datos de producción, demanda y regalías del gas natural
Equipo de Desarrollo	Valentina Fuentes, Raúl Lozano, Angelica Grazt
Ubicación	Bucaramanga, Santander
Fecha de Inicio	Noviembre 2025
Estado del Proyecto	En Desarrollo

Este proyecto fue realizado en el marco del concurso Datos al ecosistema 2025.

Contexto y Problemática Identificada

El gas natural es un recurso estratégico para la transición energética de Colombia. Sin embargo, la gestión de la información crítica para la toma de decisiones en este sector enfrenta barreras significativas debido a la fragmentación de los datos.

Situación Actual

Actualmente, los datos relacionados con la producción, demanda y regalías del gas natural en Colombia se encuentran dispersos entre diferentes entidades:

- **Producción:** Gestionada por el Ministerio de Minas y Energía (MME) - Hidrocarburos
- **Demand y Proyecciones:** Gestionada por la Unidad de Planeación Minero-Energética (UPME)
- **Liquidación de Regalías:** Gestionada por la ANH y disponible en Datos Abiertos Colombia (Socrata)

El Problema Central

No existe un punto único de consulta. Los analistas del sector, los entes de control y la ciudadanía en general deben visitar múltiples portales web, lidiar con archivos en formatos heterogéneos (Excel, CSV, consultas a diferentes APIs) y realizar cruces manuales complejos. Esta opacidad retrasa significativamente la formulación de políticas públicas basadas en evidencia y dificulta la veeduría ciudadana.

Impacto de la Problemática

- Retraso en la formulación de políticas públicas basadas en datos
- Dificultad para realizar análisis integrados entre producción, demanda y regalías
- Falta de transparencia institucional en el sector minero-energético
- Limitaciones en la veeduría ciudadana y control social

La Solución: SIMGN

SIMGN es una plataforma tecnológica unificada que actúa como un "Punto Único de Verdad" para el sector del gas natural en Colombia. Su objetivo principal es automatizar la recolección, estandarización, integración y visualización de datos dispersos, permitiendo monitorear la cadena de valor en tiempo real.

Objetivos de la Solución

- **Unificar:** Crear un punto único de acceso a datos públicos de producción, demanda y regalías
- **Automatizar:** Implementar procesos ETL que eliminan la necesidad de cruces manuales
- **Transparencia:** Democratizar el acceso a información crítica del sector
- **Evidencia:** Facilitar la toma de decisiones basada en datos confiables

Beneficios Esperados

- Análisis integrado de producción vs. demanda vs. regalías
- Identificación de brechas entre oferta y demanda
- Visualización territorial del desempeño por departamentos
- Acceso 24/7 a datos públicos estandarizados
- Descarga de datasets en formatos reutilizables (CSV, Excel)

Arquitectura Técnica y Características Principales

La solución se basa en un flujo de datos robusto y automatizado, diseñado para transformar datos crudos en conocimiento accionable, utilizando un stack tecnológico moderno (Python 3.10+ para backend/ETL y Node.js 18+ para frontend).

Componentes Principales del Flujo

ETL Automatizado (El Núcleo)

Se implementó un motor de Extracción, Transformación y Carga desarrollado en Python que:

- Se conecta a las fuentes oficiales (MME, ANH, UPME)
- Extrae datos de formatos dispersos (Excel, CSV, APIs)
- Limpia y estandariza en una estructura coherente
- Valida integridad y calidad de datos

Almacenamiento Centralizado

Los datos procesados se almacenan en una base de datos unificada (SQLite), escalable a cualquier otra base de datos relacional más estable.

Salidas y Visualización (Outputs)

- **Dashboard Interactivo:** Tablero de control web con cruce de variables
- **Inteligencia Territorial:** Mapas interactivos de indicadores por departamentos
- **Datos Abiertos:** Exportación en CSV y Excel estandarizados
- **API REST:** Acceso público e interoperabilidad

Requerimientos Funcionales

Proceso ETL

ID	Requisito	Descripción
RF-01	Extracción de Datos	Extrae automáticamente de 3 fuentes: datos.gov.co, minenergia.gov.co, upme.gov.co. Soporta CSV, Excel, JSON, APIs REST con frecuencia configurable.
RF-02	Transformación de Datos	Normaliza formatos de fecha, estandariza nombres de campos, convierte unidades, valida tipos, elimina duplicados, trata valores nulos y calcula métricas derivadas.
RF-03	Validación de Calidad	Valida rangos de valores, completitud de datos obligatorios, consistencia entre datasets, identifica outliers y registra errores.
RF-04	Carga y Almacenamiento	Carga incremental, actualización de datos existentes, versionado de datos y registro de auditoría completo.

Generación de Archivos de Salida

ID	Requisito	Descripción
RF-06	Exportación de Datos	Genera archivos Excel (.xlsx) y CSV (.csv) con datos consolidados de producción, demanda y regalías, incluyendo metadatos.
RF-07	Estructura del Archivo	Incluye: ID único, fecha/periodo, entidad territorial, campo de producción, tipo de dato, valor numérico, unidad de medida, fuente y fecha de actualización.

Dashboard Interactivo

ID	Requisito	Descripción
RF-08	Visualización de Producción	Series temporales, producción por campo y entidad territorial, comparativos históricos, mapa geográfico con filtros por fecha y campo.
RF-09	Visualización de Demanda	Series temporales, demanda por región, proyecciones vs. real, estacionalidad, comparativos años sobre año.
RF-10	Visualización de Regalías	Regalías por campo y entidad territorial, series temporales, top campos, distribución territorial con múltiples filtros.
RF-11	Análisis Integrado	Relación producción vs. regalías, balance oferta-demanda, indicadores de eficiencia, análisis de tendencias y detección de anomalías.
RF-12	Filtros y Segmentación	Rango de fechas, períodos (mensual/trimestral/anual), entidad territorial, campo de producción, tipo de dato, operador.

Gestión de Usuarios y Acceso

ID	Requisito	Descripción
RF-15	Acceso Público	Dashboard de libre acceso sin registro. Interfaz web responsive. URL pública y permanente.

Requerimientos No Funcionales

Disponibilidad y Confiabilidad

- Disponibilidad mínima del 95%
- Ventanas de mantenimiento programadas
- Operación con datos parciales si una fuente falla

Seguridad

- Comunicaciones mediante HTTPS
- Prevención de inyección SQL
- Validación de entradas

Usabilidad

- Responsive design

Conformidad y Estándares

- Cumplimiento con políticas de datos abiertos de Colombia
- Metadatos según estándares internacionales
- Licencia Creative Commons u otra clara
- Cumplimiento Ley 1712 de 2014 (Transparencia)
- Normatividad del sector minero-energético

Requerimientos de Datos

Fuentes de Datos

Fuente	Entidad	URL	Formato	Campos Clave
Consolidación de Regalías	datos.gov.co	https://www.datos.gov.co/d/j7js-yk74	CSV/API	Campo, Periodo, Valor regalías, Entidad territorial
Declaración de Producción de Gas Natural	minenergia.gov.co	https://www.minenergia.gov.co/es/misional/hidrocarburos/functionamiento-del-sector/gas-natural/	Descarga/API	Campo, Periodo, Volumen, Operador
Proyección de la Demanda de Gas Natural	upme.gov.co	https://www.upme.gov.co/simec/planeacion-energetica/proyencion_de_demandas/	Excel/API	Periodo, Región, Demanda proyectada, Real

Stack Tecnológico y Arquitectura

Tecnologías Propuestas

Capa	Tecnología	Herramientas/Librerías
Backend / ETL	Python 3.10+	pandas, requests, BeautifulSoup4
Frontend / Dashboard	Node.js 18+	React, Streamlit, Bootstrap
Base de Datos	SQLite	SQLite
Hosting/Infraestructura	Cloud (AWS)	GitHub Actions
Versionado	Git	GitHub

Flujo de Arquitectura

El sistema sigue una arquitectura modular clara:

```
[Fuentes de Datos] → [ETL en Python] → [Base de Datos] → [API REST] →
[Dashboard Web] → [Exportación CSV/Excel]
```

Componentes del Sistema

- **Módulo ETL:** Scripts Python automatizados para extracción, transformación y carga
- **Base de Datos:** SQLite centralizada con esquema optimizado y es escalable a cualquier otra base de datos relacional más estable.
- **API REST:** Endpoints para consultas y acceso programático
- **Dashboard:** Aplicación web interactiva con visualizaciones en tiempo real
- **Exportador:** Generador automático de archivos Excel/CSV consolidados

Impacto y Sostenibilidad

Descripción del Impacto

SIMGN transforma la gestión del sector minero-energético al correlacionar automáticamente datos de producción, demanda y regalías. Al unificar esta información dispersa, democratizamos su acceso y fortalecemos la transparencia institucional, permitiendo detectar patrones de mercado tan pronto las entidades publican la información.

Categorías de Impacto

Impacto Económico

Facilita el seguimiento de regalías y balance del mercado de gas, permitiendo a tomadores de decisión identificar oportunidades y riesgos en tiempo real.

Impacto Social

Garantiza la transparencia y el derecho ciudadano de acceso a datos abiertos para veeduría, permitiendo que la sociedad civil monitoree las dinámicas del sector.

Impacto Institucional/Gobernanza

Moderniza la gestión pública al unificar fuentes dispersas, permitiendo a las entidades una toma de decisiones eficiente y basada en evidencia.

Indicadores de Éxito

- Integración exitosa de las 3 fuentes de datos
- Dashboard desplegado y accesible 24/7
- Datos actualizados automáticamente
- Documentación completa al 100%
- Código en repositorio público

Plan de Implementación

Fases del Proyecto

Fase 1: Análisis y Diseño

- Exploración detallada de fuentes de datos
- Diseño de arquitectura del sistema
- Diseño del esquema de base de datos

Fase 2: Desarrollo ETL

- Implementación de extractores de datos
- Implementación de transformaciones y validaciones

Fase 3: Desarrollo Dashboard

- Implementación de visualizaciones principales
- Desarrollo de interactividad y filtros
- Integración con datos y API

Fase 5: Documentación y Entrega

- Documentación técnica completa
- Preparación de presentación para jurado
- Despliegue final en entorno de producción

Conclusiones

El Sistema de Información y Monitoreo de Gas Natural (SIMGN) representa una solución integral e innovadora para uno de los principales retos identificados por el Ministerio de Minas y Energía: la fragmentación de datos críticos para el sector de hidrocarburos en Colombia.

A través de su arquitectura modular, procesos ETL automatizados y una interfaz intuitiva, SIMGN logra transformar datos dispersos en información accesible y confiable. Este sistema no solo facilita la toma de decisiones basada en evidencia, sino que también promueve la transparencia institucional y fortalece el acceso ciudadano a datos abiertos.

El proyecto está completamente alineado con los objetivos del concurso Datos al Ecosistema 2025 y representa un compromiso genuino con la modernización de la gestión pública colombiana mediante el aprovechamiento estratégico de tecnologías abiertas y estándares de datos abiertos.

Con SIMGN, la transparencia en el sector del gas natural pasa de ser un ideal a una realidad accesible, tangible y utilizable para todos los actores interesados: desde el analista especializado hasta el ciudadano común que desea ejercer su derecho de acceso a información pública.

Documento elaborado: noviembre 2025