Metodología de Recolección de Datos

Resumen

Este documento describe cómo se recopilan, verifican y actualizan los datos del ecosistema de financiamiento verde colombiano que alimentan la plataforma Ruta Verde. Es el documento de referencia para auditores, donantes y socios que deseen verificar cualquier dato del sistema.

Fuentes Primarias

Institución	Tipo Fuente	URL	Última Verificación
Bancóldex	Página web + comunicados de prensa	bancoldex.com	Junio 2026
Findeter	Página web + informes anuales	findeter.gov.co	Junio 2026
Bancolombia	Informe Integrado 2024 + web	grupobancolombia.com/informe integrado/2024	Junio 2026
Davivienda	Informe de Sostenibilidad 2024	davivienda.com/sostenibilidad	Junio 2026
BBVA Colombia	Informe No-Financiero 2024	bbva.com.co/responsabilidad-corporativa	Junio 2026
Banco de Bogotá	Comunicados + web sostenibilidad	bancodebogota.com/sostenibilidad	Junio 2026
Banco Agrario	Informe de Gestión y Sostenibilidad 2024	bancoagrario.gov.co	Junio 2026
Bancamía	Página web productos	bancamia.com.co	Junio 2026
Banco W	Página web + comunicados bono social	bancow.com.co	Junio 2026
BID/IDB	Portal de proyectos BID	iadb.org/en/project/CO-L1289	Junio 2026
IFC	Comunicados de prensa	ifc.org	Junio 2026
Banco de la República	IBR rates vía Grupo Aval/BanRep	banrep.gov.co	2026-06-23

Método de Extracción

Importación inicial (xlsx)

El dataset base de 51 productos fue construido manualmente a partir de páginas web bancarias, informes de sostenibilidad 2024 y comunicados de prensa. Cada campo fue verificado contra la fuente primaria antes de la importación.

Pipeline automatizado semanal

El pipeline se ejecuta cada domingo a las 23:00 via systemd timer en el VPS CTH. Pasos:

1. Scraper HTTP visita páginas de productos de 6 bancos (Bancóldex, Findeter, Bancolombia, Davivienda, BBVA, Banco de Bogotá)
2. Extrae señales de cambio (variaciones en tasa IBR, disponibilidad de línea)
3. Compara contra los valores actuales en la DB MariaDB
4. Detecta diffs ? genera lista de cambios pendientes de aprobación
5. Envía digest semanal por Telegram al equipo CTH
6. Genera los tres JSON tier (admin, partners, investors) para el dashboard

Normalización de tasas

Las tasas se normalizan a **puntos base (BPS) sobre IBR overnight**. IBR como referencia: 10.51% (2026-06-23, fuente: Banco de Bogotá Internacional / Banco de la República). Las tasas no comparables en BPS (concesional, leasing, garantía) se almacenan como NULL en `rate_normalized_spread`.

Cadena de Trazabilidad

1. **Dashboard muestra el dato** ? e.g. "BCO-01 tasa: IBR + margen, descuento sostenible"
2. **DB almacena la fuente** ? tabla `data_sources`: URL, fecha de acceso, método de extracción
3. **Wiki documenta el método** ? esta página + Página 4 (Registro de Fuentes)

Frecuencia de Verificación

- **Semanal:** Pipeline automatizado (scrapers)
- **Trimestral:** Spot-check de 10 productos contra páginas web bancarias
- **Anual:** Refresh competitivo completo

Revisión #1
Creado 2026-06-24 23:05:41 UTC por Angelica Diaz
Actualizado 2026-06-24 23:05:41 UTC por Angelica Diaz