

Catálogo DIPRES (Raw)

Table of contents

1	Convención de nombres	1
2	Cómo acceder a las descargas raw	2
2.1	Trazabilidad raw → bronze → silver	2
3	Capa bronze	3
4	Unidades y naturaleza de montos	3
5	Ejecución Total — Nivel nacional	4
6	Proyecto de Ley de Presupuestos — Nivel nacional	4
6.1	Ley de Presupuestos — Nivel nacional	5
7	Metadatos generados (data/meta/dipres_files_metadata.csv)	6
7.1	Glosario de programas (data/meta/diccionarios/dipres_programas.csv)	6
7.2	Series COFOG derivadas (scripts/build_cofog_series.py)	7
8	Capa Silver normalizada	7
9	Automatización de descargas	7
10	Pendientes y diagnóstico automatizado	8
11	Próximos pasos sugeridos	8

1 Convención de nombres

dipres_ejecucion_total_{anio}-{periodo}-{moneda}.{formato}
dipres_proyecto_ley_{anio}_{slug}.{formato}

- {anio}: año fiscal (ej. 2022, 2024, 2025).
- {periodo}: nombre de mes (enero, febrero, ...) o trimestre (primer_trimestre, ...).
- {moneda}: pesos o dolares (cuando aplica).

- `{slug}`: título resumido del documento (ej. `proyecto_de_ley_de_presupuestos_nivel_nacional_pesos_definiciones_estrategicas_proyecto_de_ley_de_presupuestos_2025`).
- `{formato}`: `csv`, `xml`, `xls`, `xlsx`.

Los binarios se almacenan en `data/raw/dipres/` (ignorados en Git). El manifest con metadatos está en `data/meta/dipres_files_metadata.csv` e incluye tamaño, hash SHA-256, URL de origen, autor, institución y fecha de recuperación. El glosario de nombres para partidas/capítulos/programas se materializa en `data/meta/diccionarios/dipres_programas.csv`, generado automáticamente desde los anexos HTML que DIPRES publica por programa.

2 Cómo acceder a las descargas raw

- Cada dataset del catálogo (`dipres_ejecucion_total`, `dipres_proyecto_ley`, `dipres_ley_presupuesto` y `dipres_presupuestos`) expone en la landing un bloque “Activos raw” con conteos por año/formato/moneda y un `details` que consulta `/api/0.6.9/datasets/{slug}/raw-assets`.
- Las tablas normalizadas pueden consultarse dinámicamente vía `/api/0.6.9/data/bronze.dipres_presupuestos` (incluye filtros, paginación y linaje hacia archivos raw).
- El endpoint devuelve metadatos por archivo (año, período, variante, nivel, moneda, formato, tamaño en bytes, hash, URL de origen) y construye automáticamente el enlace `/downloads/{ruta}` para descargar el binario almacenado en `data/raw/....`
- El manifest completo sigue disponible en `/downloads/data/meta/dipres_files_metadata.csv` y puede filtrarse con herramientas externas si se requiere un batch.

2.1 Trazabilidad raw → bronze → silver

1. **Raw**: binario original en `data/raw/dipres/....`. Ejemplo: `data/raw/dipres/ley_presupuesto/2009/dipres_ley_presupuesto_2009_xls` (expuesto vía `/api/0.6.9/datasets/dipres_ley_presupuesto/raw-assets`).
2. **Bronze**: normalización CSV (`DipresBudgetBronze`)
 - `data/bronze/dipres_presupuestos_totales.csv`
 - `data/bronze/dipres_presupuestos_nodos.csv`
 - Declarado en `data/meta/catalog/dipres_presupuestos.yaml`.
 - Operaciones: `[alias trimestre→mes (marzo, junio, septiembre, diciembre)]`, `[multiplicar por 1000]`, `[conversión a CLP con tipo de cambio promedio anual]`, `[clasificación naturaleza ingreso/gasto según subtítulo]`, `[excluir Partida 50 (Tesoro Público)]`, `[enriquecer nombres de partida/capítulo/programa con el glosario dipres_programas.csv]`.
3. **Silver**: tablas `presupuestos`, `presupuesto_partidas`, ..., `presupuesto_asignaciones` en Postgres (migración `202510090002_add_presupuestos.py`).
4. **Gold / API**: endpoints `/api/{ver}/presupuestos` y dashboards (`/dashboard/presupuesto`).

Actualmente sólo almacenamos montos **nominales** (moneda corriente). La columna `unidad` indica el factor aplicado (`factor_multiplicado=1000`) y la moneda destino (CLP). Los montos reales/deflactados quedan pendientes para una fase posterior.

3 Capa bronze

El pipeline `DipresBudgetBronze` genera dos archivos listos para la capa bronze:

- `data/bronze/dipres_presupuestos_totales.csv`: una fila por presupuesto (dataset + período + moneda + variante) con los montos agregados; las columnas `_mo` mantienen el valor publicado por DIPRES (miles de pesos/dólares) y las columnas `_clp` ya están convertidas a pesos chilenos corrientes.
- `data/bronze/dipres_presupuestos_nodos.csv`: detalle jerárquico (partida → capítulo → programa → subtítulo → ítem → asignación) con los montos en moneda original (miles) y CLP (pesos), códigos parentales y naturaleza contable.

Ambos archivos incluyen el slug del presupuesto, la variante (`base`, `ley_regular_incendios`, `fet_covid19`, etc.), el nivel de origen (`nacional`, `partida`, ...), la ruta del archivo fuente y la fecha de recuperación desde el portal DIPRES. La portada (/) consume esta metadata desde `data/meta/catalog/dipres_presupuestos.yaml` para exponer descargas, esquema y una muestra de cada tabla bronze.

- **Clasificación de transferencias**: el neteo intragubernamental utiliza el diccionario `data/meta/diccionarios/dipres_transfer_destinos.csv`, que etiqueta cada subtítulo/ítem de las transferencias corrientes (24) y de capital (33) según el sector receptor (`sector_privado`, `gobierno_central`, `entidades_publicas`, `sector_externo`, etc.). Los registros marcados como `intragobierno=true` se descuentan explícitamente de los totales netos (`monto_ejecutado_neto_*`). El diccionario puede extenderse añadiendo nuevas filas con comodines (*) para ítems o asignaciones específicas. Además, se aplica una heurística por denominación que reclasifica automáticamente los montos dirigidos a municipalidades o programas de gratuidad en educación superior fuera del perímetro intragobierno, alineando la serie con el total oficial de DIPRES.
- **Variantes regulares vs extraordinarias**: además de los agregados jerárquicos (`partida`, `capitulo`, `programa`, `subtitulo`), la carpeta `tmp/` y el payload `aggregates["variant"]` entregan una tabla resumen por año y variante (`base`, `fet_covid19`, `incendios`, ...) con los montos brutos, ajustes intragobierno y netos. Esto permite separar rápidamente el gasto extraordinario (por decretos FET/incendios) del presupuesto base.

4 Unidades y naturaleza de montos

- DIPRES publica los montos en **miles** de pesos o dólares. El pipeline normaliza todos los valores multiplicando por 1 000 antes de exponerlos en la API. Los campos `monto_*_clp` ya están en pesos chilenos (CLP) y los campos `monto_*_mo` reflejan la moneda original, igualmente convertida desde miles a unidades.
- Cada nodo de la jerarquía incorpora un bloque `unidad` que detalla el factor aplicado y las monedas involucradas, de modo que no se realicen conversiones adicionales por error.
- El campo `naturaleza` indica la clasificación contable del registro:
 - `ingreso`: subtítulos 01 a 15.
 - `gasto`: subtítulos 21 a 35.
 - `mixto`: agregados con ingresos y gastos.

- desconocido: filas sin subtítulo o con códigos no catalogados.
- Para convertir montos a términos reales y compararlos con el PIB, se dispone del paquete `ref_bcch` (`data/meta/catalog/ref_bcch.yaml`) que contiene:
 - `data/ref/bcch/ipc_mensual.csv` — IPC general histórico (base 2018=100) descargado vía `scripts/fetch_bcch_series.py`.
 - `data/ref/bcch/pib_trimestral_real_2018.csv` — PIB real encadenado trimestral (base 2018=100, CLP de 2018). Estas series permiten construir deflatores y razones monto/PIB en la capa analítica (silver/gold).

5 Ejecución Total — Nivel nacional

Cobertura descargada automáticamente (CSV + XML en pesos y dólares):

Año	Mensual disponibles	Mensual faltantes	Trimestral disponibles	Trimestral faltantes
2025	enero, febrero, abril, mayo, julio, agosto	marzo, junio, septiembre, octubre, noviembre, diciembre	primer_trimestre, segundo_trimestre	tercer_trimestre, cuarto_trimestre
2024	enero, febrero, abril, mayo, julio, agosto, octubre, noviembre, diciembre	marzo, junio, septiembre	primer_trimestre, segundo_trimestre, tercer_trimestre	cuarto_trimestre
2023	enero, febrero, abril, mayo, julio, agosto, octubre, noviembre, diciembre	marzo, junio, septiembre	primer_trimestre, segundo_trimestre, tercer_trimestre	cuarto_trimestre

Observaciones:

- **2023:** la pestaña “Nacional” se carga vía AJAX y no expone los enlaces en el HTML inicial; quedó pendiente automatizar ese fetch.
- **2021:** el patrón `w3-multipropertyvalues-25910-{pvid}.html` devuelve 404. Se requiere investigar endpoints históricos o repositorios alternativos de DIPRES.
- Sólo se recolectó el nivel nacional; las pestañas de Partidas/Capítulos/Programas también cargan contenido asincrónico y no están incluidas en esta pasada.
- Los antiguos “primer/segundo/tercer/cuarto trimestre” se normalizan como **marzo, junio, septiembre y diciembre**. El manifest y la API sólo exponen meses.

6 Proyecto de Ley de Presupuestos — Nivel nacional

Descargas nacionales disponibles (CSV/XML en pesos):

Año	Instrumento	Nivel	Moneda	Formato	URL
2025	Proyecto Nacional	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-346717_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-346717_doc_xml.xml
2024	Proyecto Nacional	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-320810_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-320810_doc_xml.xml
2023	Proyecto Nacional	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-288938_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-288938_doc_xml.xml
2022	Proyecto Nacional	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-244740_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-244740_doc_xml.xml
2021	Proyecto Nacional	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-208646_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-208646_doc_xml.xml
2020	Proyecto Nacional	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-193253_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-193253_doc_xml.xml
2019	Proyecto Nacional	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-178466_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-178466_doc_xml.xml
2018	Proyecto Nacional	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-166184_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-166184_doc_xml.xml
2017	Proyecto Nacional	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-150135_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-150135_doc_xml.xml
2016	Proyecto Nacional	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-136005_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-136005_doc_xml.xml
2015	Proyecto Nacional	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-122591_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-122591_doc_xml.xml
2014	Proyecto Nacional	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-114118_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-114118_doc_xml.xml
2013	Proyecto Nacional	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-94669_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-94669_doc_xml.xml

Complementos (`definiciones_estrategicas_*`, `indicadores_de_desempeno_*`) se mantienen para 2022–2025 (XLSX). Las pestañas Partida/Capítulo/Programa siguen cargándose vía AJAX.

2010 y años anteriores retornan 404 en `w3-multipropertyvalues-15168-{pvid}.html`.
Hace falta revisar el archivo histórico de DIPRES o el repositorio 594/598.

6.1 Ley de Presupuestos — Nivel nacional

Año	Instrumento	Nivel	Moneda	Formato	URL
2025	Ley	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-354806_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-354806_doc_xml.xml
2024	Ley	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-325081_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-325081_doc_xml.xml
2023	Ley	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-299462_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-299462_doc_xml.xml
2022	Ley	Nacional	Pesos	CSV	https://www.dipres.gob.cl/597/articles-257515_doc_csv.csv ·
				XML	https://www.dipres.gob.cl/597/articles-257515_doc_xml.xml

Año	Instrumento	Nivel	Moneda	Formato	URL
2021	Ley	Nacional	Pesos	CSV, XML	https://www.dipres.gob.cl/597/articles-213730_doc_csv.csv · https://www.dipres.gob.cl/597/articles-213730_doc_xml.xml
2020	Ley	Nacional	Pesos	CSV, XML	https://www.dipres.gob.cl/597/articles-198054_doc_csv.csv · https://www.dipres.gob.cl/597/articles-198054_doc_xml.xml
2019	Ley	Nacional	Pesos	CSV, XML	https://www.dipres.gob.cl/597/articles-183868_doc_csv.csv · https://www.dipres.gob.cl/597/articles-183868_doc_xml.xml
2018	Ley	Nacional	Pesos	CSV, XML	https://www.dipres.gob.cl/597/articles-168762_doc_csv.csv · https://www.dipres.gob.cl/597/articles-168762_doc_xml.xml
2017	Ley	Nacional	Pesos	CSV, XML	https://www.dipres.gob.cl/597/articles-155375_doc_csv.csv · https://www.dipres.gob.cl/597/articles-155375_doc_xml.xml

No se encontraron enlaces públicos para 2016 y años anteriores; el portal responde 404.
Se requiere revisar el sitio legacy o el archivo PDF oficial.

7 Metadatos generados (data/meta/dipres_files_metadata.csv)

Para cada archivo se registra:

- `dataset`, `instrument_type`
- año, periodo, tipo de periodo (`month`, `quarter`, `annual`)
- moneda (`pesos`, `dolares`, `sin_moneda`)
- formato (`csv`, `xml`, `xls`, `xlsx`)
- nombre de archivo y ruta local
- URL DIPRES y título oficial
- `retrieved_at` (UTC), autor/institución
- tamaño en bytes y hash SHA-256
- estado (`downloaded` si se bajó en esta corrida, `existing` si ya estaba)

7.1 Glosario de programas (data/meta/diccionarios/dipres_programas.csv)

- Generado automáticamente durante el pipeline `DipresBudgetBronze` a partir de los anexos `*_presupuesto_programa_*` publicados por DIPRES.
- Llave primaria: `partida_codigo`, `capitulo_codigo`, `programa_codigo` (normalizados a dos dígitos).
- Para cada jerarquía se elige la descripción más informativa; si DIPRES entrega variantes en distintos años el pipeline deja un warning y retiene la etiqueta más extensa.
- Disponible en `/downloads/meta/diccionarios/dipres_programas.csv` para reutilizar la misma taxonomía en terceros procesos o dashboards.
- La misma corrida publica `/downloads/meta/diccionarios/dipres_cofog_programas.csv`, puente contra el Anexo 4 (COFOG adaptado) para poder agrupar por función/subfunción.

7.2 Series COFOG derivadas (scripts/build_cofog_series.py)

- Lee `data/bronze/dipres_presupuestos_nodos.csv` (nivel programa) y cruza con el glosario COFOG.
- Entrega agregados anuales con montos nominales, montos reales base 2018, participación del gasto total y % del PIB por subfunción (`data/analytics/cofog_subfuncion_anual.csv`) y función (`data/analytics/cofog_funcion_anual.csv`).
- Calcula automáticamente deflatores (IPC base 2018), totales de presupuesto y PIB real para exponer montos reales, % del gasto público y % del PIB, insumo directo para los gráficos presupuestarios.

8 Capa Silver normalizada

Los presupuestos descargados se normalizan en la capa *silver* según el esquema `data/meta/schema/dipres_presupuestos`.

- El manifiesto raw (`data/meta/dipres_files_metadata.csv`) se cruza con tasas de cambio (`data/ref/dipres_exchange_rates.csv`).
- Cada archivo genera jerarquía Presupuesto → Partida → Capítulo → Programa → Subtítulo → Ítem → Asignación con montos en moneda original y CLP.
- Conversión a CLP usa promedio anual publicado por el Banco Central (ver `data/ref/dipres_exchange_rates`).
- Diccionario de subtítulos: `data/meta/diccionarios/dipres_subtitulos.yaml`.
- Dataset catalogado en `data/meta/catalog/dipres_presupuestos.yaml` (incluye referencias al glosario `data/meta/diccionarios/dipres_programas.csv`).

El API expone esta jerarquía bajo `/api/{version}/presupuestos` con filtros, ordenamiento y paginación. Ver `docs/api_catalog.md` para rutas detalladas.

Para análisis comparativo se disponibilizan las vistas `vw_presupuesto_medidas` y `vw_presupuesto_variacion_anual` además del endpoint `/api/{version}/presupuestos/stats` (agregados, filtros por nivel y variaciones interanuales).

9 Automatización de descargas

- Script: `python scripts/download_dipres_history.py`
 - Acepta filtros vía `DIPRES_YEARS=2025,2024` o rangos `DIPRES_MIN_YEAR=2018 DIPRES_MAX_YEAR=2025` para evitar cargas masivas.
 - Reintenta sólo el nivel nacional (`*_nacional_*`) y normaliza los nombres de archivo a `<dataset>_<año>-<periodo>_nacional_<moneda>.<ext>`.
 - Las descargas quedan en `data/raw/dipres/<dataset>/<año>/` y el manifest se actualiza con hash SHA-256 y tamaño en bytes.

10 Pendientes y diagnóstico automatizado

Ejecuta `PYTHONPATH=. python scripts/check_dipres_coverage.py` (o consulta `/api/0.6.9/status`) para obtener el resumen actualizado de cobertura. Resultado al 2025-10-12:

- **Ejecución Total 2003-2014:** DIPRES sólo publica informes al cierre de trimestre. Los meses disponibles son **marzo, junio, septiembre y diciembre**. El primer trimestre 2003 no está disponible en el portal.
- **Ejecución Total 2015-2024:** cobertura mensual completa (enero → diciembre) en pesos y dólares.
- **Ejecución Total 2025:** se cuenta con enero-agosto. Los informes de septiembre a diciembre aún no existen (el ejercicio sigue en curso) y se etiquetan como **upcoming**.
- **Ejecución Total 2023:** descargada íntegramente (12 informes CSV/XML por moneda) y normalizada en `data/bronze/dipres_presupuestos_{totales,nodos}.csv`.
- **Proyecto de Ley 2013-2026 y Ley de Presupuesto 2003-2025:** disponibles en CSV/XML nacional.

Actualmente `pending_downloads` queda vacío; el script sólo reportará nuevos pendientes cuando aparezcan informes adicionales o cambie la cobertura DIPRES.

11 Próximos pasos sugeridos

1. Automatizar la descarga para años 2023 y anteriores resolviendo los endpoints dinámicos de DIPRES.
2. Incorporar pestañas adicionales (Ley Regular, Partidas, GORES) y formatos complementarios según necesidades del proyecto.
3. Programar un job periódico que verifique periodos faltantes y actualice el manifest, informando en el status page si persisten pendientes.