

Package: tampaBayOpenData (via r-universe)

October 24, 2026

Title Discover and Retrieve Tampa Bay Public Open Data

Version 0.1.0

Description Provides tools for discovering and retrieving public 'ArcGIS' datasets across the Tampa Bay region. Search a selected catalog of checked datasets or live publisher portals, and load records into 'R'. The six configured publishers are the City of Tampa, City of St. Petersburg, City of Clearwater, Hillsborough County, Pinellas County, and Tampa Bay Regional Planning Council. Discover layers in these and other public 'ArcGIS' portals and retrieve them by their service URLs. A publisher registry configures discovery, while a separate dataset registry identifies layers checked by package maintainers. Filter catalogs by source, geometry, topic, category, and modification date. Retrieve queryable layers with validated pagination, bounded previews, server-side filtering, typed tabular results, optional spatial objects, and source provenance. Save large downloads as local chunks that can be validated and resumed. Public source pages include the City of Tampa GeoHub <<https://city-tampa.opendata.arcgis.com/>>, St. Petersburg 'ArcGIS' portal <<https://csp.maps.arcgis.com/home/index.html>>, Clearwater 'ArcGIS' portal <<https://cityofclearwater.maps.arcgis.com/home/index.html>>, Hillsborough County GeoHub <<https://hcfl.gov/about-hillsborough/open-data-and-gis/geohub>>, Pinellas County Enterprise GIS <<https://new-pinellas-egis.opendata.arcgis.com/>>, and Tampa Bay Regional Planning Council GeoHub <<https://opendata-tbrpc.hub.arcgis.com/>>.

License MIT + file LICENSE

URL <https://jaclenga.github.io/tampaBayOpenData/>,
<https://github.com/Jaclenga/tampaBayOpenData>

BugReports <https://github.com/Jaclenga/tampaBayOpenData/issues>

Encoding UTF-8

```
Depends R (>= 4.1.0)
Imports httr2 (>= 1.2.3), jsonlite (>= 1.6.0), tibble (>= 3.0.0)
Suggests ggplot2, knitr, rmarkdown, sf (>= 1.0.0), testthat (>= 3.2.1)
VignetteBuilder knitr
Config/testthat/edition 3
Roxygen list(markdown = TRUE)
Config/roxygen2/version 8.1.0
Config/pak/sysreqs libssl-dev
Repository https://jaclenga.r-universe.dev
Date/Publication 2026-10-24 12:42:21 UTC
RemoteUrl https://github.com/jaclenga/tampabayopendata
RemoteRef HEAD
RemoteSha a70e6d73b8595e37442abb0e251e46fd2e85b202
```

Contents

tbod_check_catalog	2
tbod_check_schema	4
tbod_dataset_info	5
tbod_discover	6
tbod_download_arcgis_layer	7
tbod_download_dataset	9
tbod_get_arcgis_layer	11
tbod_get_capital_projects	13
tbod_get_dataset	13
tbod_get_development_cases	16
tbod_get_permits	17
tbod_list_datasets	17
tbod_list_portals	19
tbod_provenance	20
tbod_schema	20
tbod_search_datasets	21
Index	24

tbod_check_catalog	<i>Check the current health of checked catalog layers</i>
--------------------	---

Description

Checks current public ArcGIS endpoints and saved schemas for selected entries in the bundled checked catalog. It contacts publisher services and returns a report without changing the catalog or saving files. A healthy result is a point-in-time metadata check, not a measure of record freshness.

Usage

```
tbod_check_catalog(
  jurisdiction = "all",
  ids = NULL,
  timeout = 10,
  total_timeout = 1800,
  pause = 0.25,
  probe_query = FALSE
)
```

Arguments

<code>jurisdiction</code>	One checked-catalog jurisdiction code or "all".
<code>ids</code>	Optional nonempty vector of unique checked dataset IDs within jurisdiction; NULL checks all selected entries.
<code>timeout</code>	Per-HTTP-attempt timeout in seconds.
<code>total_timeout</code>	Maximum elapsed seconds for the complete check, including requests, retries, and pauses. Defaults to 1,800; Inf disables the overall deadline. Entries not reached are reported as unknown.
<code>pause</code>	Delay from 0 to 30 seconds between entries to limit request rate; defaults to 0.25. With <code>probe_query = TRUE</code> , also pauses between layer metadata and the count probe.
<code>probe_query</code>	Run a zero-match count query after usable metadata to verify Query capability. Defaults to FALSE for the existing metadata-only behavior. Query probes do not retrieve features.

Details

Each entry starts one layer-metadata request. The client probes the service root only after a missing-layer response (HTTP or ArcGIS 404/410, or an explicit layer-not-found error) to distinguish a missing layer from a missing service. Transient failures remain unknown after bounded retries; entries not reached before `total_timeout` are also unknown. The default retry limit is three additional attempts per retryable failure sequence. Set `options(tampaBayOpenData.max_retries = 3L)` to change it (whole number from 0 to 10).

`healthy` means the layer metadata is valid, Query is advertised, and the schema matches the saved snapshot. `schema_changed` includes both compatible and incompatible drift; inspect `schema_status` and `schema_issues`. Confirmed schema drift remains `schema_changed` if Query capability is unreported. A missing object-ID field is reported as schema drift when a baseline exists; the details also identify unsupported record retrieval. With no saved baseline, `schema_status` is unavailable and `status` is unknown. `unsupported` includes layers that cannot be queried and access failures. By default the check does not retrieve features or run count and object-ID queries. A layer can advertise Query yet fail those operations during retrieval. Set `probe_query = TRUE` to run a zero-match count request after valid layer metadata; it reads no features and records whether the actual Query operation succeeds. A confirmed query failure is degraded, while a transient probe failure is unknown unless independently confirmed schema drift already makes the result degraded. The result cannot guarantee future service availability.

Value

A tibble with one row per selected entry and publisher, jurisdiction, dataset_id, endpoint, checked_at (UTC POSIXct), status (healthy, unavailable, schema_changed, unsupported, or unknown), failure_type, http_status, service_status, and layer_status (available, unavailable, or unknown), schema_status (unchanged, compatible, incompatible, or unavailable), schema_issues (a list of issue tibbles), and details. health_state summarizes status as healthy, degraded, unavailable, or unknown. response_time_ms is elapsed request time including retries and a missing-layer service-root probe, excluding deliberate pauses; it is missing if no request was made. query_status is not_checked, verified, failed, or unknown; query_failure_type, query_http_status, and query_details explain a failed or uncertain probe. http_status is missing when no metadata HTTP response was received.

Examples

```
if (identical(Sys.getenv("TAMPA_OPEN_DATA_LIVE"), "true")) {
  report <- tbod_check_catalog(ids = "construction-permits")
  report[, c("dataset_id", "status", "schema_status", "details")]
}
```

tbod_check_schema	<i>Check a current ArcGIS schema against an expected schema</i>
-------------------	---

Description

The checked catalog supplies an expected snapshot automatically. For a discovered descriptor or direct ArcGIS URL, pass expected explicitly to compare it with a previously saved [tbod_schema\(\)](#) result. Otherwise the result is untracked and contains the current schema. Added fields and changed aliases are compatible drift. Removed or renamed fields, changed field types, object-ID fields, geometry or CRS, and moved or missing endpoints are incompatible drift. This inspection function returns a report, including missing endpoints; checked retrieval warns for compatible drift and errors before querying for incompatible drift.

Usage

```
tbod_check_schema(
  id,
  jurisdiction = NULL,
  expected = NULL,
  timeout = 30,
  total_timeout = 60,
  portal = NULL
)
```

Arguments

id	Checked dataset ID, discovered descriptor, stable ArcGIS item ID, or public HTTPS ArcGIS layer URL.
----	---

jurisdiction	Optional checked-catalog jurisdiction. NULL resolves a unique bundled ID and preserves a discovery row's source. Direct URLs do not accept a jurisdiction override.
expected	Optional schema list with endpoint, fields, geometry_type, spatial_reference, and object_id_field. A saved <code>tbod_schema()</code> result can be passed directly.
timeout	Per-request timeout in seconds.
total_timeout	Overall operation budget in seconds; Inf disables it.
portal	Optional public ArcGIS sharing REST portal URL for a stable arcgis:<item-id>:<layer-id> identifier. It does not apply to other IDs.

Value

A `tbod_schema_check` list with status (unchanged, compatible, incompatible, or untracked), a tibble of issues with category, severity, field, expected, and actual columns, and expected and current schemas. An incompatible missing endpoint has no current schema.

Examples

```
if (identical(Sys.getenv("TAMPA_OPEN_DATA_LIVE"), "true")) {
  report <- tbod_check_schema("construction-permits")
  report$status
  report$issues
}
```

tbod_dataset_info	<i>Inspect a dataset and its source</i>
-------------------	---

Description

Accepts a bundled ID, one-row discovery result, stable ArcGIS ID, or direct public HTTPS layer URL. A checked ID reads its bundled record offline; a stable ArcGIS ID resolves the current portal item. `refresh = TRUE` retrieves current fields, aliases, types, geometry, CRS, query capabilities, and editing metadata. The registry's verified date records an endpoint check, not record freshness. Use `tbod_schema()` to inspect the schema or `tbod_check_schema()` to compare it with the checked snapshot.

Usage

```
tbod_dataset_info(
  id,
  jurisdiction = NULL,
  refresh = FALSE,
  timeout = 30,
  total_timeout = 60,
  portal = NULL
)
```

Arguments

<code>id</code>	Bundled ID, one-row discovery result, stable ArcGIS ID, or public HTTPS ArcGIS layer URL.
<code>jurisdiction</code>	Optional bundled catalog jurisdiction. NULL resolves a unique bundled ID and preserves a discovery result's jurisdiction. Direct URLs do not accept a jurisdiction override.
<code>refresh</code>	Whether to request current ArcGIS layer metadata.
<code>timeout</code>	Per-request timeout in seconds.
<code>total_timeout</code>	Maximum elapsed seconds for inspection, including metadata requests and retries. Defaults to 60; Inf disables it.
<code>portal</code>	Optional ArcGIS sharing REST portal URL when <code>id</code> is a stable <code>arcgis:<item-id>:<layer-id></code> identifier.

Value

A named list with the dataset ID, title, publisher, jurisdiction, source and service URLs, scope, and terms. With `refresh = TRUE`, it also contains `fields` (a tibble of source names, types, and aliases), `metadata` (raw ArcGIS layer metadata), and UTC inspected `_at`.

Examples

```
info <- tbod_dataset_info("construction-permits")
info[c("id", "title", "publisher", "source_url")]
```

<code>tbod_discover</code>	<i>Discover datasets from an ArcGIS portal</i>
----------------------------	--

Description

Searches a public ArcGIS portal or lists layers from a FeatureServer or MapServer service or layer URL. Pass a result row to `tbod_get_dataset()` to retrieve it. Portal rows have `discovered` status; service URL rows have `not_checked` status because they have no portal item identity. Publisher services can change after discovery.

Usage

```
tbod_discover(
  portal,
  query = NULL,
  max_items = 25,
  timeout = 30,
  total_timeout = 90,
  org_id = NULL,
  publisher = NULL,
  jurisdiction = "unspecified"
)
```

Arguments

portal	Public HTTPS ArcGIS sharing REST root ending in /sharing/rest, or FeatureServer/MapServer service or layer URL.
query	Optional portal search text. For a service URL, a literal case-insensitive layer-name filter.
max_items	Maximum service items to scan in a portal or layers to inspect in a service. Defaults to 25; Inf scans all available entries. A capped portal search can omit matching layers from older items.
timeout	Per-request timeout in seconds.
total_timeout	Maximum elapsed seconds for the operation. Defaults to 90; Inf disables this limit.
org_id	Optional ArcGIS organization ID to restrict a portal search. This argument does not apply to service URLs.
publisher	Publisher label to attach to discovery rows. If NULL, a portal organization name is used when available; service URLs use an unknown-publisher label.
jurisdiction	Jurisdiction label to attach to discovery rows. Defaults to "unspecified" for a custom portal.

Details

A portal search scans at most `max_items` service items, then expands queryable layers. For a FeatureServer or MapServer URL, discovery enumerates the service's layers; query filters their names. The result reports scan completeness and item errors as attributes.

Value

A tibble of discovered ArcGIS layers. Discovery issues, scan counts, truncation, and completeness are attached as attributes; check `attr(rows, "discovery_complete")` before assuming a complete search.

Examples

```
if (identical(Sys.getenv("TAMPA_OPEN_DATA_LIVE"), "true")) {
  rows <- tbod_discover("https://www.arcgis.com/sharing/rest",
    query = "parks", max_items = 5)
}
```

tbod_download_arcgis_layer

Download a direct ArcGIS layer URL

Description

Uses `tbod_download_dataset()` with a public HTTPS ArcGIS FeatureServer or MapServer layer URL. The direct source has `not_checked` provenance and no inferred publisher identity. A saved manifest fixes the endpoint, query, fields, schema, CRS, IDs, and batch size; a resumed call rechecks them before reusing validated chunks. Successful validation writes `complete.rds` in the download directory; a resumed call clears and recreates it after matching the manifest and revalidating chunks. See `tbod_download_dataset()` for the directory, lock, and record-change rules.

Usage

```
tbod_download_arcgis_layer(
  url,
  path,
  where = "1=1",
  fields = NULL,
  spatial = FALSE,
  out_sr = NULL,
  page_size = 1000,
  query = list(),
  timeout = 30,
  total_timeout = Inf,
  resume = TRUE
)
```

Arguments

<code>url</code>	Public HTTPS ArcGIS layer URL.
<code>path</code>	Local directory for the manifest, checkpoints, and chunk files.
<code>where</code>	ArcGIS SQL WHERE clause using source field names; defaults to "1=1". Inspect names with <code>tbod_dataset_info(url, refresh = TRUE)</code> .
<code>fields</code>	Character vector of exact source field names, or NULL or "*" for all nongeometry attributes. Object IDs remain in chunk provenance even if the column was not selected.
<code>spatial</code>	Write sf chunks; requires the optional sf package and a layer with supported geometry and a known CRS.
<code>out_sr</code>	Optional positive ArcGIS/EPSG output reference for server projection; requires <code>spatial = TRUE</code> . NULL keeps the native CRS.
<code>page_size</code>	Positive integer batch size capped at 1,000 and the service maximum. The effective size must stay fixed when resuming.
<code>query</code>	Named list of advanced ArcGIS filters such as geometry, spatialRel, and time. Response format, field selection, aggregation, geometry simplification, and pagination parameters are protected.
<code>timeout</code>	Per-request timeout in seconds. Transient failures may be retried according to <code>options(tampaBayOpenData.max_retries = 3L)</code> .
<code>total_timeout</code>	Overall operation deadline in seconds. The default Inf allows a long download; completed chunks remain for a later resume if a finite deadline expires.
<code>resume</code>	Reuse a matching saved download and skip validated chunks.

Value

A named list with the download path, ordered chunk files, matched and returned row counts, completeness, and summary provenance.

Examples

```
if (identical(Sys.getenv("TAMPA_OPEN_DATA_LIVE"), "true")) {
  url <- tbod_dataset_info("construction-permits")$source_url
  tbod_download_arcgis_layer(url, tempfile("permit-url-download-"),
    where = "OBJECTID <= 10", page_size = 5)
}
```

tbod_download_dataset *Download a dataset into resumable chunks*

Description

Accepts a bundled ID, discovery row, stable ArcGIS ID, or direct ArcGIS layer URL. A custom Enterprise portal can be provided with stable IDs. Layers in the checked catalog are compared with expected schemas before downloading; compatible drift warns and incompatible drift stops the download.

Usage

```
tbod_download_dataset(
  id,
  path,
  jurisdiction = NULL,
  where = "1=1",
  fields = NULL,
  spatial = FALSE,
  out_sr = NULL,
  page_size = 1000,
  query = list(),
  timeout = 30,
  total_timeout = Inf,
  resume = TRUE,
  portal = NULL
)
```

Arguments

id	Bundled ID, one-row discovery result, stable ArcGIS ID, or direct public HTTPS ArcGIS layer URL.
path	Local directory for the manifest, checkpoints, and chunk files.

jurisdiction	Optional bundled catalog jurisdiction. NULL resolves a unique bundled ID and preserves a discovery result's jurisdiction. Direct URLs do not accept a jurisdiction override.
where	ArcGIS SQL WHERE clause using source field names; defaults to "1=1". Inspect names with <code>tbod_dataset_info(id, refresh = TRUE)</code> .
fields	Character vector of exact source field names, or NULL or "*" for all nongeometry attributes. Object IDs remain in chunk provenance even if the column was not selected.
spatial	Write sf chunks; requires the optional sf package and a layer with supported geometry and a known CRS.
out_sr	Optional positive ArcGIS/EPSG output reference for server projection; requires <code>spatial = TRUE</code> . NULL keeps the native CRS.
page_size	Positive integer batch size capped at 1,000 and the service maximum. The effective size must stay fixed when resuming.
query	Named list of advanced ArcGIS filters such as geometry, spatialRel, and time. Response format, field selection, aggregation, geometry simplification, and pagination parameters are protected.
timeout	Per-request timeout in seconds. Transient failures may be retried according to <code>options(tampaBayOpenData.max_retries = 3L)</code> .
total_timeout	Overall operation deadline in seconds. The default Inf allows a long download; completed chunks remain for a later resume if a finite deadline expires.
resume	Reuse a matching saved download and skip validated chunks.
portal	Optional ArcGIS sharing REST portal URL when id is a stable <code>arcgis:<item-id>:<layer-id></code> identifier.

Details

The client retrieves the matching object-ID manifest when records exist, then writes bounded RDS chunks to a caller-selected directory. Each chunk is a tibble or sf object with its own provenance; read files one at a time with `readRDS()`. Feature pages are not accumulated in memory, but the manifest itself is subject to the one-million-record limit.

A saved manifest fixes the endpoint, query, fields, schema, CRS metadata, matching IDs, and batch size. With `resume = TRUE`, current metadata and IDs are rechecked before saved chunks are used. Changed options, schema, CRS, or membership cause an error. Completed chunks must match their checkpoint checksums, expected IDs, columns, and provenance. Temporary files from interrupted writes are ignored. After all chunks have been validated, the client atomically writes `complete.rds` in path. No marker is written before a new download finishes or after an interrupted call. After confirming the saved manifest matches, a resumed call clears the marker before revalidating saved chunks, then writes it again only after successful completion.

The directory must be empty on the first call. Unrelated files and existing downloads with `resume = FALSE` are rejected. An exclusive lock prevents concurrent writers. If a terminated process leaves `.lock`, confirm no writer remains, remove that empty directory, and resume. Record attributes can change between chunks or resumed calls even when IDs and schema remain stable; this is not a frozen snapshot. The returned summary records whether the complete matching manifest was

downloaded. Transient connection failures and timeouts, HTTP 429, 500, 502, 503, and 504 responses, ArcGIS error objects with those codes inside HTTP 200 responses, and a generic transient ArcGIS query error are retried. Set `options(tampaBayOpenData.max_retries = 3L)` to control additional retries per failure sequence (integer 0 through 10). HTTP failures and ArcGIS errors in successful HTTP responses use separate bounded sequences; mixed failures can make more than four HTTP attempts overall. A valid Retry-After delay or HTTP date is honored if at most 30 seconds; a longer delay stops retrying rather than causing an early retry. Otherwise retries use exponential backoff with jitter. Permanent HTTP and ArcGIS errors fail without retrying. An interrupted call does not return a complete download summary; resume with the same path to validate and continue saved chunks.

Value

A named list with path, ordered chunk files, matched_rows, returned_rows, complete, and summary source provenance. Each RDS chunk has its own provenance; `tbod_provenance(chunk)$download` includes its chunk number and object IDs. With no matches, files is empty and summary integrity is count-only.

Examples

```
if (identical(Sys.getenv("TAMPA_OPEN_DATA_LIVE"), "true")) {
  saved <- tbod_download_dataset("construction-permits",
    tempfile("permit-download-"), where = "OBJECTID <= 10", page_size = 5)
  if (length(saved$files)) tbod_provenance(readRDS(saved$files[[1L]]))
}
```

`tbod_get_arcgis_layer` *Retrieve a direct ArcGIS layer URL*

Description

Calls `tbod_get_dataset()` with a public HTTPS ArcGIS FeatureServer or MapServer layer URL. Direct layers have not_checked provenance and no inferred publisher or portal item identity.

Usage

```
tbod_get_arcgis_layer(
  url,
  where = "1=1",
  fields = NULL,
  spatial = FALSE,
  out_sr = NULL,
  order_by = NULL,
  limit = Inf,
  page_size = NULL,
  query = list(),
  timeout = 30,
  total_timeout = 120,
```

```
    integrity = "auto"
  )
```

Arguments

<code>url</code>	Public HTTPS ArcGIS layer URL.
<code>where</code>	ArcGIS SQL WHERE clause using source field names; defaults to "1=1". Inspect names with <code>tbod_dataset_info(url, refresh = TRUE)</code> .
<code>fields</code>	Character vector of exact source field names, or NULL or "*" for all nongeometry attributes. An object-ID field is requested internally for integrity checks and omitted from the result if unselected.
<code>spatial</code>	Return an <code>sf</code> object. Requires the optional <code>sf</code> package and supported geometry with a reliably determined CRS.
<code>out_sr</code>	Optional positive ArcGIS/EPSG spatial reference ID for server projection, such as 4326. Requires <code>spatial = TRUE</code> ; NULL keeps the native CRS. The response must confirm a requested projection.
<code>order_by</code>	Source fields with ASC or DESC, such as <code>c("LASTUPDATE DESC", "RECORD_ID ASC")</code> . Requires server ordering and offset pagination; an object-ID tie breaker is added automatically.
<code>limit</code>	Maximum records. Inf retrieves all; zero returns a typed empty result. A deliberate subset is <code>complete = FALSE</code> in <code>tbod_provenance()</code> when more matching records exist.
<code>page_size</code>	Optional positive integer batch size, capped at 1,000 and the service maximum.
<code>query</code>	Named list of advanced ArcGIS filters: <code>geometry</code> , <code>geometryType</code> , <code>inSR</code> , <code>spatialRel</code> , <code>relationParam</code> , <code>time</code> , <code>distance</code> , <code>units</code> , <code>gdbVersion</code> , <code>historicMoment</code> , <code>sqlFormat</code> , and <code>datumTransformation</code> . Values can be scalars or JSON-style lists. Response format, field selection, aggregation, geometry simplification, and pagination parameters are protected.
<code>timeout</code>	Per-request timeout in seconds. Transient failures may be retried according to <code>options(tampaBayOpenData.max_retries = 3L)</code> .
<code>total_timeout</code>	Overall operation budget in seconds. Defaults to 120; Inf disables it. Requests and retry waits use the remaining budget.
<code>integrity</code>	"auto" (default) uses bounded preview validation when a finite limit is at most 1,000, more than 10,000 records match, and the layer supports ordering and pagination. The returned IDs are checked against the original filter. If advertised offsets fail or repeat, an automatic preview can restart with a full manifest and object-ID batches. Explicit <code>order_by</code> requests do not use this fallback. Full retrieval always uses the full object-ID manifest when matches exist; "full" requires it even for a preview. A zero limit or zero matching count requests only the count. Provenance records the method used.

Value

A tibble or `sf` object with source, completeness, and retrieval metadata attached. Read it with `tbod_provenance()`.

Examples

```
if (identical(Sys.getenv("TAMPA_OPEN_DATA_LIVE"), "true")) {
  url <- tbod_dataset_info("construction-permits")$source_url
  data <- tbod_get_arcgis_layer(url, limit = 10)
  tbod_provenance(data)$validation_status
}
```

tbod_get_capital_projects

Retrieve checked City of Tampa capital projects

Description

Calls `tbod_get_dataset()` with the checked City of Tampa capital-projects catalog ID to retrieve current publisher records. The City service publishes records marked PUBLIC; the package preserves that source scope.

Usage

```
tbod_get_capital_projects(...)
```

Arguments

... Arguments passed to `tbod_get_dataset()` except id.

Value

A tibble or sf object with retrieval provenance attached.

Examples

```
if (identical(Sys.getenv("TAMPA_OPEN_DATA_LIVE"), "true")) {
  tbod_get_capital_projects(limit = 10)
}
```

tbod_get_dataset

Retrieve a checked, discovered, or direct ArcGIS dataset

Description

Accepts a checked catalog ID, a one-row discovery result, a stable `arcgis:<item-id>:<layer-id>` identifier, or a public HTTPS ArcGIS layer URL ending in `/FeatureServer/<id>` or `/MapServer/<id>`. For a stable ArcGIS ID from an Enterprise portal, supply that portal's sharing REST URL in portal. Direct URLs use the same retrieval engine and carry `not_checked` provenance. For a layer in the checked catalog, retrieval compares current metadata with its expected schema, warns for compatible drift, and errors before querying for incompatible drift. Use `tbod_check_schema()` to inspect the differences.

Usage

```
tbod_get_dataset(
  id,
  jurisdiction = NULL,
  where = "1=1",
  fields = NULL,
  spatial = FALSE,
  out_sr = NULL,
  order_by = NULL,
  limit = Inf,
  page_size = NULL,
  query = list(),
  timeout = 30,
  total_timeout = 120,
  integrity = "auto",
  portal = NULL
)
```

Arguments

<code>id</code>	Bundled ID, one-row discovery result, stable <code>arcgis:<item-id>:<layer-id></code> identifier, or public HTTPS ArcGIS layer URL ending in <code>/FeatureServer/<id></code> or <code>/MapServer/<id></code> .
<code>jurisdiction</code>	Optional checked-catalog jurisdiction. NULL resolves a unique checked ID across the bundled catalog and preserves the source jurisdiction of a discovery result. Direct URLs use "unspecified" and do not accept a jurisdiction override.
<code>where</code>	ArcGIS SQL WHERE clause using source field names; defaults to "1=1" for all records. Inspect names with <code>tbod_dataset_info(id, refresh = TRUE)</code> .
<code>fields</code>	Character vector of exact source field names, or NULL or "*" for all nongeometry attributes. An object-ID field is requested internally for integrity checks and omitted from the result if unselected.
<code>spatial</code>	Return an <code>sf</code> object. Requires the optional <code>sf</code> package and supported geometry with a reliably determined CRS.
<code>out_sr</code>	Optional positive ArcGIS/EPSSG spatial reference ID for server projection, such as 4326. Requires <code>spatial = TRUE</code> ; NULL keeps the native CRS. The response must confirm a requested projection.
<code>order_by</code>	Source fields with ASC or DESC, such as <code>c("LASTUPDATE DESC", "RECORD_ID ASC")</code> . Requires server ordering and offset pagination; an object-ID tie breaker is added automatically.
<code>limit</code>	Maximum records. Inf retrieves all; zero returns a typed empty result. A deliberate subset is <code>complete = FALSE</code> in <code>tbod_provenance()</code> when more matching records exist.
<code>page_size</code>	Optional positive integer batch size, capped at 1,000 and the service maximum.
<code>query</code>	Named list of advanced ArcGIS filters: <code>geometry</code> , <code>geometryType</code> , <code>inSR</code> , <code>spatialRel</code> , <code>relationParam</code> , <code>time</code> , <code>distance</code> , <code>units</code> , <code>gdbVersion</code> , <code>historicMoment</code> , <code>sqlFormat</code> ,

	and datumTransformation. Values can be scalars or JSON-style lists. Response format, field selection, aggregation, geometry simplification, and pagination parameters are protected.
timeout	Per-request timeout in seconds. Transient failures may be retried according to <code>options(tampaBayOpenData.max_retries = 3L)</code> .
total_timeout	Overall operation budget in seconds. Defaults to 120; Inf disables it. Requests and retry waits use the remaining budget.
integrity	"auto" (default) uses bounded preview validation when a finite limit is at most 1,000, more than 10,000 records match, and the layer supports ordering and pagination. The returned IDs are checked against the original filter. If advertised offsets fail or repeat, an automatic preview can restart with a full manifest and object-ID batches. Explicit <code>order_by</code> requests do not use this fallback. Full retrieval always uses the full object-ID manifest when matches exist; "full" requires it even for a preview. A zero limit or zero matching count requests only the count. Provenance records the method used.
portal	Optional public HTTPS ArcGIS sharing REST portal URL for a stable <code>arcgis:<item-id>:<layer-id></code> identifier. Omit to use the global ArcGIS item lookup. This argument does not apply to other input types.

Details

Retrieval requests the current layer schema and matching count. Full requests with matches also retrieve and validate an object-ID manifest. A small preview of a large queryable layer can validate only its returned IDs against the original filter. Truncated or short object-ID batches are split into smaller requests. Persistently missing or duplicate records, truncated manifests, unexpected responses, and upstream errors fail explicitly. Full retrieval requires a complete manifest and must be narrowed by filters if more than one million records match. Use `tbod_download_dataset()` to save validated chunks to disk.

Source field names, strings, coded values, and date-looking strings are preserved. ArcGIS epoch-millisecond dates become UTC POSIXct and date-only fields become Date. Dates declared to have an unknown timezone remain raw milliseconds with a warning; UTC editor tracking dates can still convert. Integer values remain integers when safe, larger numeric values become doubles, and exact big integers can be represented as character values. NULL or absent attributes become typed missing values. Spatial results use two-dimensional XY coordinates and exclude Z and M values.

The package does not cache live data. Record membership changes detected during retrieval cause an error, but attribute changes cannot be frozen. Accepted responses are limited to 50 MiB and 64 nested JSON containers; requests do not follow redirects. Native curl, GDAL, and PROJ parsing remains subject to the installed libraries' behavior.

Transient connection failures and timeouts, HTTP 429, 500, 502, 503, and 504 responses, ArcGIS error objects with those codes inside HTTP 200 responses, and a generic transient ArcGIS query error are retried. The default is three additional retries per failure sequence; set the integer option `tampaBayOpenData.max_retries` from 0 to 10 to change it. HTTP failures and ArcGIS errors in successful HTTP responses have separate bounded sequences, so mixed failures may use more than four HTTP attempts overall. Retries use exponential backoff with jitter or a valid `Retry-After` delay or HTTP date. A requested delay longer than 30 seconds stops retrying instead of retrying before the service permits it. Invalid queries, authentication failures, unavailable layers, and unsupported operations fail without retries. `total_timeout` may stop attempts earlier.

Value

A tibble, or an sf object when `spatial = TRUE`. `source`, `dataset_id`, `jurisdiction`, and `retrieved_at` attributes record the source and retrieval. Read the complete record with [tbod_provenance\(\)](#).

Examples

```
if (identical(Sys.getenv("TAMPA_OPEN_DATA_LIVE"), "true")) {
  data <- tbod_get_dataset("construction-permits", limit = 5)
  tbod_provenance(data)$source_url
}
```

```
tbod_get_development_cases
```

Retrieve checked City of Tampa development cases

Description

Calls [tbod_get_dataset\(\)](#) with the checked City of Tampa development-cases catalog ID to retrieve current publisher records. It covers active entitlement locations, not all historical development cases.

Usage

```
tbod_get_development_cases(...)
```

Arguments

... Arguments passed to [tbod_get_dataset\(\)](#) except `id`.

Value

A tibble or sf object with retrieval provenance attached.

Examples

```
if (identical(Sys.getenv("TAMPA_OPEN_DATA_LIVE"), "true")) {
  tbod_get_development_cases(limit = 10)
}
```

tbod_get_permits	<i>Retrieve checked City of Tampa permit locations</i>
------------------	--

Description

Calls `tbod_get_dataset()` with the checked City of Tampa construction-permits catalog ID to retrieve current publisher records. It reflects the City's active permit viewer; the source is not a complete historical archive of every permit.

Usage

```
tbod_get_permits(...)
```

Arguments

... Arguments passed to `tbod_get_dataset()` except id.

Value

A tibble or sf object with retrieval provenance attached.

Examples

```
if (identical(Sys.getenv("TAMPA_OPEN_DATA_LIVE"), "true")) {
  tbod_get_permits(limit = 10)
}
```

tbod_list_datasets	<i>List datasets in the bundled catalog</i>
--------------------	---

Description

By default, scans the configured Tampa Bay ArcGIS portals and combines their layers with the bundled checked catalog. Use `source = "checked"` to list checked entries offline. Live scans may be incomplete; see Details for the result attributes. Use `tbod_discover()` to scan another portal.

Usage

```
tbod_list_datasets(
  jurisdiction = "all",
  source = "all",
  portals = "all",
  max_items = 25,
  timeout = 30,
  total_timeout = 90,
  publisher = NULL,
```

```

validation_status = NULL,
spatial = NULL,
topic = NULL,
category = NULL,
modified_after = NULL,
modified_before = NULL
)

```

Arguments

jurisdiction	Jurisdiction code or vector of codes from <code>tbod_list_portals()</code> , or "all" (default). Filters checked and live results and narrows the portal scan.
source	"all" (default) combines live discovery with checked entries; "live" returns portal layers with checked matches marked; "checked" reads only the offline registry.
portals	Configured publisher IDs from <code>tbod_list_portals()</code> , or "all". This selects live portals and does not filter checked entries.
max_items	Maximum number of service items scanned per selected portal. Defaults to 25; Inf follows all search pages. An item can expand to multiple layers, so this does not bound the row count.
timeout	Per-request timeout in seconds.
total_timeout	Overall operation budget, including retries and live discovery. Defaults to 90 seconds; Inf disables the deadline.
publisher	Publisher name or vector, matched exactly without regard to case. Narrows both checked results and the live portal scan.
validation_status	"checked", "discovered", or a vector of both; NULL includes both. source = "live" can include checked matches.
spatial	TRUE selects layers with declared geometry, FALSE selects nonspatial tables, and NULL includes both and unknown geometry.
topic	Word or vector matched as case-insensitive literal substrings in tags and categories; NULL disables the filter.
category	Category or vector matched, ignoring case, against full source paths or their final labels.
modified_after, modified_before	Inclusive modification bounds. A Date or YYYY-MM-DD string includes its full UTC day; POSIXct specifies an instant. Unknown modification times are excluded when a bound is set. This is an item or checked snapshot time, not record freshness.

Details

Values within a filter are combined with OR; different filters and a search query are combined with AND. Status, geometry, topic, category, and date filters apply after the bounded live scan and checked overlay. A capped scan can miss older matching items. Use `max_items = Inf` with an appropriate time budget to follow all available index pages, and inspect `attr(result, "discovery_complete")`.

This flag describes the selected public portal index scan, not every publisher service. Live results also carry `discovery_issues`, `discovery_failed_portals`, `discovery_scanned_items`, and `discovery_truncated_portals` attributes. If every selected portal fails, `source = "all"` returns the checked entries with failure details in these attributes; unknown scan counts are NA.

Value

A tibble with stable IDs, service and layer locations, portal and item IDs when available, and `validation_status` of "checked" or "discovered". Live results include scan-completeness and issue attributes described in Details.

Examples

```
tbod_list_datasets(source = "checked")
```

tbod_list_portals	<i>List bundled ArcGIS publisher configurations</i>
-------------------	---

Description

Returns the package's Tampa Bay publisher configurations without a network request. Each entry provides an ArcGIS organization ID and sharing REST root. Its verified date checks publisher identity, not every layer's contents or reuse terms. To scan another portal, pass its sharing REST URL to [tbod_discover\(\)](#).

Usage

```
tbod_list_portals()
```

Value

A tibble with `id` (the discovery selector), `root`, `org_id`, `publisher`, `jurisdiction`, `website`, `verified` (a Date), `metadata_url`, and `evidence_url`. Use `id` in the `portals` argument of [tbod_list_datasets\(\)](#) or [tbod_search_datasets\(\)](#).

Examples

```
tbod_list_portals()
```

tbod_provenance	<i>Read retrieval provenance</i>
-----------------	----------------------------------

Description

Reads the source, query, completeness, and schema information attached to a tabular or spatial result from `tbod_get_dataset()`, or a chunk saved by `tbod_download_dataset()`. The record includes publisher, source and query URLs, UTC retrieval time, returned and matching row counts, and integrity status. `validation_status` is checked for bundled entries, discovered for live portal results, and `not_checked` for direct layer URLs. Item ID, portal, and original metadata are included when available. A successful full retrieval checks object-ID membership, but upstream attributes are not a transactionally frozen snapshot. R transformations can discard attributes; save provenance before transforming or exporting the result.

Usage

`tbod_provenance(x)`

Arguments

`x` A retrieval result or a chunk read from `tbod_download_dataset()` or `tbod_download_arcgis_layer()`

Value

The attached source list, including `dataset_id`, `publisher`, `jurisdiction`, `source_url`, `query_endpoint`, `validation_status`, UTC `started_at` and `retrieved_at`, submitted query, `matched_rows`, `returned_rows`, `complete`, and `integrity`. Download chunks also include download with the chunk number and object IDs.

Examples

```
if (identical(Sys.getenv("TAMPA_OPEN_DATA_LIVE"), "true")) {
  permits <- tbod_get_permits(limit = 10)
  tbod_provenance(permits)$source_url
}
```

tbod_schema	<i>Inspect a dataset schema</i>
-------------	---------------------------------

Description

For a checked dataset, the default returns the bundled schema snapshot without a network request. `refresh = TRUE` requests current ArcGIS layer metadata. A direct ArcGIS layer URL or discovered descriptor has no bundled expected schema and requires `refresh = TRUE`. Call `tbod_check_schema()` to compare a current service against its expected schema. A stable ArcGIS item ID from an Enterprise portal can be resolved with `portal`.

Usage

```
tbod_schema(
  id,
  jurisdiction = NULL,
  refresh = FALSE,
  timeout = 30,
  total_timeout = 60,
  portal = NULL
)
```

Arguments

id	Checked dataset ID, discovered descriptor, stable ArcGIS item ID, or public HTTPS ArcGIS layer URL.
jurisdiction	Optional checked-catalog jurisdiction. NULL resolves a unique bundled ID and preserves a discovery row's source. Direct URLs do not accept a jurisdiction override.
refresh	Request current layer metadata when TRUE.
timeout	Per-request timeout in seconds.
total_timeout	Overall operation budget in seconds; Inf disables it.
portal	Optional public ArcGIS sharing REST portal URL for a stable arcgis:<item-id>:<layer-id> identifier. It does not apply to other IDs.

Value

A list with endpoint, a tibble of field names, types and aliases, geometry type, spatial reference, object-ID field, and dataset ID.

Examples

```
schema <- tbod_schema("construction-permits")
head(schema$fields[, c("name", "type")])
```

tbod_search_datasets *Search checked and live ArcGIS datasets*

Description

Searches checked Tampa Bay entries and, when requested, configured live ArcGIS portals. The checked catalog is a selected, offline registry; live discovery searches each publisher's public portal index. Use [tbod_discover\(\)](#) to search another ArcGIS portal or a service URL.

Usage

```

tbod_search_datasets(
  query,
  jurisdiction = "all",
  source = "all",
  portals = "all",
  max_items = 25,
  timeout = 30,
  total_timeout = 90,
  publisher = NULL,
  validation_status = NULL,
  spatial = NULL,
  topic = NULL,
  category = NULL,
  modified_after = NULL,
  modified_before = NULL,
  limit = Inf
)

```

Arguments

query	A single search string. Checked matching treats punctuation literally and ignores case. Separate words must all match. An empty or whitespace-only query lists datasets.
jurisdiction	Jurisdiction code or vector of codes from tbod_list_portals() , or "all" (default). Filters checked and live results and narrows the portal scan.
source	"all" (default) combines live discovery with checked entries; "live" returns portal layers with checked matches marked; "checked" reads only the offline registry.
portals	Configured publisher IDs from tbod_list_portals() , or "all". This selects live portals and does not filter checked entries.
max_items	Maximum number of service items scanned per selected portal. Defaults to 25; Inf follows all search pages. An item can expand to multiple layers, so this does not bound the row count.
timeout	Per-request timeout in seconds.
total_timeout	Overall operation budget, including retries and live discovery. Defaults to 90 seconds; Inf disables the deadline.
publisher	Publisher name or vector, matched exactly without regard to case. Narrows both checked results and the live portal scan.
validation_status	"checked", "discovered", or a vector of both; NULL includes both. source = "live" can include checked matches.
spatial	TRUE selects layers with declared geometry, FALSE selects nonspatial tables, and NULL includes both and unknown geometry.
topic	Word or vector matched as case-insensitive literal substrings in tags and categories; NULL disables the filter.

category	Category or vector matched, ignoring case, against full source paths or their final labels.
modified_after, modified_before	Inclusive modification bounds. A Date or YYYY-MM-DD string includes its full UTC day; POSIXct specifies an instant. Unknown modification times are excluded when a bound is set. This is an item or checked snapshot time, not record freshness.
limit	Maximum result rows returned after filtering and ranking. Defaults to Inf; use a nonnegative whole number, including 0, or Inf. This does not cap the number of live service items scanned; see <code>max_items</code> .

Details

Values within a filter are combined with OR; different filters and the search query are combined with AND. Publisher names are exact, case-insensitive filters; use `tbod_list_portals()` to inspect them. A publisher with no match returns no rows.

Checked entries search ID, title, description, publisher, category, and tags. Exact title or ID matches rank first, then title or ID prefixes, substrings, and matches across title or ID words. Tags, categories, and publisher matches follow; description-only matches rank after those. Within a rank, each word scores title, ID, tags or categories, publisher, and description in that order; title, ID, and source URL then break ties consistently. If no checked entry matches the literal query, a spelling fallback may return close checked entries. Query words must be alphanumeric, and words shorter than four characters have no typo tolerance. The fallback does not change live portal search terms.

Live items come from the portal index, which can search fields not exposed in returned rows. Available row metadata helps rank them, but an indexed hit is retained even without a local text match. Layers found both live and in the checked catalog appear once with "checked" status. With `source = "all"`, checked matches remain available even when a portal does not return them. `limit` applies after discovery, overlay, filtering, and ranking. Even `limit = 0` still performs the selected live scan and preserves its diagnostics.

Portal searches include public Feature Service and Map Service items and expand supported, queryable layers. A private, unindexed, stale, or unavailable item may be absent. `max_items` caps service items per portal; item errors and capped scans can leave returned results incomplete, and a timeout can stop the operation. Check `attr(result, "discovery_complete")` and the issue, failed-portal, scanned-item, and truncated-portal attributes described in `tbod_list_datasets()`. Even a complete scan covers the public portal index, not every publisher service.

Value

A tibble with the same columns and live discovery attributes as `tbod_list_datasets()`. A checked-only search requires no network request.

Examples

```
tbod_search_datasets("permit", source = "checked")
```

Index

`readRDS()`, [10](#)

`tbod_check_catalog`, [2](#)
`tbod_check_schema`, [4](#)
`tbod_check_schema()`, [5](#), [13](#), [20](#)
`tbod_dataset_info`, [5](#)
`tbod_discover`, [6](#)
`tbod_discover()`, [17](#), [19](#), [21](#)
`tbod_download_arcgis_layer`, [7](#)
`tbod_download_arcgis_layer()`, [20](#)
`tbod_download_dataset`, [9](#)
`tbod_download_dataset()`, [8](#), [15](#), [20](#)
`tbod_get_arcgis_layer`, [11](#)
`tbod_get_capital_projects`, [13](#)
`tbod_get_dataset`, [13](#)
`tbod_get_dataset()`, [6](#), [11](#), [13](#), [16](#), [17](#), [20](#)
`tbod_get_development_cases`, [16](#)
`tbod_get_permits`, [17](#)
`tbod_list_datasets`, [17](#)
`tbod_list_datasets()`, [19](#), [23](#)
`tbod_list_portals`, [19](#)
`tbod_list_portals()`, [18](#), [22](#), [23](#)
`tbod_provenance`, [20](#)
`tbod_provenance()`, [12](#), [14](#), [16](#)
`tbod_schema`, [20](#)
`tbod_schema()`, [4](#), [5](#)
`tbod_search_datasets`, [21](#)
`tbod_search_datasets()`, [19](#)