

# Base de Dados da OMC: Controvérsias 1995–2022 – Metodologia de Construção

Estruturando 251 relatórios do sistema de solução de controvérsias via extração com Claude API e análise em R

## Alan Leal

Economista, Theta Lab

### Sumário Executivo

Este documento descreve a metodologia de construção da base de dados de controvérsias da Organização Mundial do Comércio (OMC) para o período 1995–2022. A base cobre 251 relatórios de painel e/ou Órgão de Apelação (OA) adotados pelo Órgão de Solução de Controvérsias (DSB), estruturados a partir dos *One-Page Case Summaries* publicados pela OMC. A extração foi realizada via Claude API (Anthropic), que processou cada resumo em prompt estruturado, gerando variáveis sobre partes, produto, acordo, resultado e duração.

## ORIGEM DOS DADOS

### WTO Dispute Settlement One-Page Case Summaries

A fonte primária é a publicação oficial da OMC *Dispute Settlement: One-Page Case Summaries 1995–2022* (edição de 2023), que reúne resumos padronizados de todos os relatórios adotados pelo DSB desde o início do funcionamento do sistema de solução de controvérsias em 1995. Cada resumo, de aproximadamente uma página, contém:

- Número(s) DS e título abreviado do caso;
- Partes reclamante(s), respondente e terceiros;
- Datas de estabelecimento do painel, distribuição dos relatórios de painel e OA, e adoção pelo DSB;
- Produto e medida contestada;
- Acordos OMC invocados e resultado do caso.

A publicação cobre exclusivamente casos cujo relatório foi adotado pelo DSB. Disputas encerradas por acordo mútuo, casos com relatório circulado mas ainda não adotado à data de corte, e pedidos de consultas sem estabelecimento de painel não estão incluídos.

A publicação está disponível em formato PDF no portal da OMC: [wto.org](https://wto.org) – [Dispute Settlement One-Page Summaries 2023](#).

### Revisões de Cumprimento (Artigo 21.5 do DSU)

O Entendimento sobre Solução de Controvérsias (DSU) permite que, após a adoção de uma decisão, o membro vencedor questione a adequação das medidas de cumprimento adotadas pelo respondente. Esses procedimentos do Artigo 21.5 geram relatórios adicionais, igualmente incluídos na publicação. Na base, tais casos são identificados pelo campo `is_compliance = TRUE` e pelo campo `compliance_round`, que numera as rodadas de revisão (1, 2, 3...). As análises percentuais e win rates reportadas neste documento referem-se, salvo indicação em contrário, apenas aos **210 casos originais** (`is_compliance = FALSE`).

## EXTRAÇÃO VIA CLAUDE API

### Arquitetura do pipeline

Cada resumo de uma página foi processado individualmente pela API da Anthropic (endpoint `/v1/messages`, modelo `claude-sonnet-4-6`) com um prompt estruturado que solicitava a extração em formato JSON. O script `main_extraction.R` (incluído na distribuição deste Data Brief) implementa o pipeline completo em R, utilizando os pacotes `httr2`, `jsonlite`, `pdftools`, `dplyr` e `stringr`.

O fluxo de processamento consiste em cinco etapas:

- 1. Extração de texto do PDF:** o pacote `pdftools` converte cada página do PDF em texto bruto, preservando a estrutura de parágrafos dos resumos.
- 2. Segmentação por caso:** o texto é segmentado por número DS, isolando o bloco de texto de cada caso individualmente.
- 3. Chamada à API:** cada bloco é enviado ao Claude com instruções para retornar exclusivamente JSON com os campos definidos no schema da base.
- 4. Parsing e validação:** o JSON retornado é parseado e validado contra o schema; registros com campos obrigatórios ausentes são sinalizados para revisão manual.
- 5. Enriquecimento:** variáveis derivadas (grupos de renda, classificação Norte-Sul, duração em dias) são calculadas em R após a extração.

### Prompt de extração

O prompt enviado ao Claude seguiu o seguinte padrão:

```
prompt <- paste0(
  "Extract the following fields from this WTO dispute summary",
  "and return ONLY valid JSON, no explanation:\n\n",
  "Fields: ds_number, short_title, is_compliance,",
  "complainants_raw, respondent_raw, third_parties_raw,"
```

```
"date_panel_established,▯date_panel_report,▯date_ab_report,▯",  
"date_adopted,▯had_ab,▯product_description,▯measure_description,▯",  
"measure_type,▯sector,▯is_agricultural,▯hs2_code,▯hs4_code,▯",  
"agreements_cited,▯primary_agreement,▯n_articles_contested,▯",  
"n_articles_violation,▯outcome,▯ab_reversed_panel,▯",  
"ab_reversal_direction,▯had_subsequent_compliance,▯had_retaliation",  
"\n\nSummary▯text:\n", case_text  
)
```

O parâmetro `temperature` foi fixado em 0 para maximizar determinismo na extração. Cada chamada utilizou `max_tokens = 1000`, suficiente para cobrir o JSON de um único caso.

### Chamada à API em R

```
resp <- req |>  
  request("https://api.anthropic.com/v1/messages") |>  
  req_headers(  
    "x-api-key" = Sys.getenv("ANTHROPIC_API_KEY"),  
    "anthropic-version" = "2023-06-01",  
    "content-type" = "application/json"  
  ) |>  
  req_body_json(list(  
    model = "claude-sonnet-4-6",  
    max_tokens = 1000,  
    messages = list(list(role = "user", content = prompt))  
  )) |>  
  req_perform()  
  
json_text <- resp_body_json(resp)$content[[1]]$text  
result <- fromJSON(json_text)
```

## ESTRUTURA DA BASE PRIMÁRIA (WTO SUMMARIES)

Os campos extraídos diretamente dos resumos da OMC, sem transformação, são os seguintes:

Campo extraído	Nome na base	Descrição
Número DS	ds_number	Ex.: "DS2", "DS103,DS113"
Título abreviado	short_title	Denominação oficial OMC
Reclamante(s)	complainants_raw	Nome(s) textual(is)
Respondente	respondent_raw	País que defende a medida
Terceiros	third_parties_raw	Partes interessadas
Data estabelecimento	date_panel_established	Painel instaurado
Data relatório painel	date_panel_report	Distribuição do relatório
Data relatório OA	date_ab_report	NA se não houve OA
Data adoção DSB	date_adopted	Âncora temporal principal
Produto	product_description	Texto livre
Medida contestada	measure_description	Texto livre
Tipo de medida	measure_type	Categoria classificada
Acordos citados	agreements_cited	Lista separada por ";"
Acordo principal	primary_agreement	ADA, GATT, ASCM etc.
Resultado	outcome	complainant_wins, mixed, respondent_wins
Reversão pelo OA	ab_reversed_panel	TRUE/FALSE
Direção da reversão	ab_reversal_direction	pro_complainant, pro_respondent, both

## PRÉ-PROCESSAMENTO E VARIÁVEIS DERIVADAS

### Identificação ISO3 e enriquecimento de partes

Os nomes textuais de países extraídos pelo Claude foram convertidos para códigos ISO 3166-1 alpha-3 em duas etapas:

- **Match exato:** comparação case-insensitive entre o nome extraído e a tabela de referência de membros OMC, que inclui denominações oficiais em inglês e francês;
- **Dicionário de aliases:** casos residuais com grafias alternativas ("European Communities" → EUU, "Separate Customs Territory of Taiwan" → TWN) ou membros com representação especial são tratados por lookup estático.

Casos consolidados (múltiplos números DS na mesma linha) e disputas com múltiplos reclamantes recebem valores concatenados por ";" em `complainant_iso3`. O campo `n_complainants` registra o número de reclamantes.

### Classificação Norte-Sul

A assimetria Norte-Sul é operacionalizada pela classificação de renda do Banco Mundial (edição 2022):

Grupo WB	Critério	Label N-S
High income	RNB per capita > US\$ 13.845	North
Upper-middle income	US\$ 4.466-13.845	South
Lower-middle income	US\$ 1.136-4.465	South
Low income	< US\$ 1.136	South

A combinação do perfil de reclamante e respondente gera o campo `dispute_ns_type` com quatro categorias: `North_North`, `North_South`, `South_North` e `South_South`. Casos com múltiplos reclamantes de grupos distintos recebem `mixed` no campo `complainant_income_group`.

### Variáveis de duração

A partir das datas extraídas, foram calculadas três variáveis de duração em dias (inteiro):

Variável	Lógica de construção
<code>duration_total_days</code>	<code>date_adopted - date_panel_established</code>
<code>duration_panel_days</code>	<code>date_panel_report - date_panel_established</code>
<code>duration_ab_days</code>	<code>date_ab_report - date_panel_report</code> (somente se <code>had_ab = TRUE</code> )

### Flags booleanos derivados

- `had_ab`: TRUE se `date_ab_report` não é NA;
- `is_compliance`: TRUE se o título do caso contém "Article 21.5" ou equivalente;
- `is_agricultural`: TRUE se `hs2_code` ∈ {1...24} ou `sector = ``goods``` e `primary_agreement = ``AA```.

## ESTRUTURA DA BASE FINAL

O dataset final (`wto_disputes.csv`) é organizado em uma linha por caso (ou por rodada de cumprimento, quando `is_compliance = TRUE`), com os seguintes grupos de colunas:

Grupo	Colunas	N
Identificação	<code>ds_number</code> , <code>short_title</code> , <code>is_compliance</code> , <code>compliance_round</code>	4
Partes	<code>complainants_raw</code> , <code>respondent_raw</code> , <code>third_parties_raw</code> , <code>complainant_iso3</code> , <code>respondent_iso3</code> , <code>third_parties_iso3</code> , <code>n_complainants</code>	7
Datas	<code>date_panel_established</code> , <code>date_panel_report</code> , <code>date_ab_report</code> , <code>date_adopted</code>	4
Duração	<code>had_ab</code> , <code>duration_total_days</code> , <code>duration_panel_days</code> , <code>duration_ab_days</code>	4
Produto e medida	<code>product_description</code> , <code>measure_description</code> , <code>measure_type</code> , <code>sector</code> , <code>is_agricultural</code> , <code>hs2_code</code> , <code>hs2_description</code> , <code>hs4_code</code>	8
Acordos	<code>agreements_cited</code> , <code>primary_agreement</code> , <code>n_articles_contested</code> , <code>n_articles_violation</code>	4
Resultado	<code>outcome</code> , <code>ab_reversed_panel</code> , <code>ab_reversal_direction</code> , <code>had_subsequent_compliance</code> , <code>had_retaliation</code>	5
Assimetria N-S	<code>complainant_income_group</code> , <code>respondent_income_group</code> , <code>dispute_ns_type</code>	3
<b>Total</b>		<b>39</b>

## ACHADOS PRINCIPAIS

Os números a seguir referem-se aos **210 casos originais** (`is_compliance = FALSE`) salvo indicação em contrário.

### Taxas de resultado

O sistema favorece amplamente quem aciona: **57%** dos casos resultam em vitória do reclamante, **34%** em resultado misto e apenas **9%** em vitória do respondente. A raridade de vitórias do respondente sugere que a decisão de protocolar uma disputa já funciona como filtro de mérito – países acionam a OMC quando têm base jurídica sólida.

### Órgão de Apelação

**59%** dos casos originais chegaram ao OA. Desses, **63%** tiveram o relatório do painel revertido ou significativamente modificado – uma taxa extraordinariamente alta para um tribunal de segunda instância. A imprevisibilidade do OA é frequentemente citada como motivação central para o bloqueio norte-americano às nomeações de membros a partir de 2017.

### Assimetria Norte-Sul

A disparidade é estrutural e mensurável:

Tipo	N	Win rate recl.	Duração mediana	AB rate
North → North	74	46%	~580 dias	62%
North → South	48	<b>77%</b>	~510 dias	55%
South → North	58	55%	~620 dias	65%
South → South	22	59%	~540 dias	55%

A diferença de 22 pontos percentuais entre *North → South* (77%) e *South → North* (55%) constitui uma das maiores assimetrias estruturais documentadas em arbitragem internacional.

### Antidumping como campo de batalha dominante

Com **75 casos (30%)** do total, o Acordo Antidumping (ADA) é o mais litigado da história da OMC – quase o dobro do segundo (GATT, 50 casos). Ferro e aço (capítulos HS 72 e 73 combinados) concentram 33 disputas, tornando o setor siderúrgico o mais contestado de toda a amostra.

### Declínio estrutural do sistema

A atividade atingiu pico em 2000 (18 adoções) e declinou consistentemente a seguir, com aceleração após 2016. A paralisia do OA a partir de dezembro de 2019 – quando o quórum mínimo de três membros deixou de ser mantido – efetivamente desativou o mecanismo de apelação, deixando recursos em aberto indefinidamente e fraturando o modelo de adjudicação em dois níveis que a OMC foi construída para sustentar.

## LIMITAÇÕES E NOTAS METODOLÓGICAS

- **Escopo temporal:** a base cobre apenas relatórios adotados até dezembro de 2022. Casos com relatório circulado mas pendente de adoção, e todos os casos protocolados após 2022, estão fora do escopo.
- **Granularidade HS:** o campo `hs2_code` refere-se ao capítulo do produto *principal* da disputa. Casos que envolvem múltiplos capítulos HS têm apenas o mais relevante registrado; uma linha por capítulo não é gerada (diferente do tratamento de NCMs na base de defesa comercial brasileira – Data Brief No. 01).
- **Acordos múltiplos:** `primary_agreement` captura o acordo com maior número de artigos contestados. Quando há empate, a classificação segue hierarquia interna de relevância definida no prompt de extração. O campo `agreements_cited` preserva todos os acordos invocados.
- **Resultado vs. implementação:** `outcome` reflete o resultado jurídico formal do relatório adotado – não o grau de implementação pelo respondente. Um caso `complainant_wins` pode ter tido cumprimento deficiente; o campo `had_subsequent_compliance` sinaliza quando houve procedimento formal de verificação de cumprimento.
- **Revisões de cumprimento aninhadas:** o campo `compliance_round` permite identificar múltiplas rodadas de revisão do mesmo caso base, mas a base não encadeia automaticamente as rodadas em uma hierarquia relacional. Para análises longitudinais por caso, recomenda-se agrupar por `ds_number` após filtrar `is_compliance = FALSE` para selecionar o caso original.
- **Extração por LLM – possíveis erros residuais:** apesar da validação manual de amostra aleatória (n = 20), erros de extração pontuais não podem ser descartados. Usuários que identifiquem inconsistências são encorajados a reportar ao contato abaixo.

---

## Fontes e Referências

**World Trade Organization (2023).** *Dispute Settlement: One-Page Case Summaries 1995–2022*. Geneva: WTO Publications. Disponível em: [wto.org](https://wto.org).

**Anthropic (2024–2026).** *Claude API – Messages endpoint*. Modelo utilizado: `claude-sonnet-4-6`. Disponível em: [docs.anthropic.com](https://docs.anthropic.com).

**World Bank (2022).** *World Bank Country and Lending Groups – Income Classifications*. Washington, D.C.: World Bank. Disponível em: [datahelpdesk.worldbank.org](https://datahelpdesk.worldbank.org).

**Leal, A. (2024–2026).** *comexTL: Dados do Comércio Exterior Brasileiro no R*. Theta Lab. Disponível em: [github.com/alanleal-econ/comexTL](https://github.com/alanleal-econ/comexTL).

---

As opiniões expressas são de responsabilidade do autor. Contato: [alan.leal@thetalab.com.br](mailto:alan.leal@thetalab.com.br)