# INTERPOL DATA SCRAPING DOCUMENTATION

## 1. Goal

Extract and store all Interpol Red Notice records, including paginated data and detailed entries, into a structured database.

## 2. Process Breakdown

### Step 1: Identify API Endpoint

Used browser DevTools → Network tab to monitor requests during browsing of the Interpol Red Notices page.  
URL observed: https://www.interpol.int/en/How-we-work/Notices/View-Red-Notices

### Step 2: Extract Page-wise Links

A Python scraper was written to iterate over multiple paginated result pages, extract each Red Notice ID, and prepare detail page links.

### Step 3: Visit “View More” for Each Notice

Each Red Notice had a View More button that revealed further data such as:  
- Nationality  
- Entity ID  
- Status  
- Action Taken  
- Region

### Step 4: Combine All into CSVs

Country-wise or page-wise results were stored into individual CSV files, later zipped for upload and further use.

### Step 5: Merge and Clean Duplicates

Traversed the unzipped folder and combined all \*\_combined\_data.csv files into a master CSV.  
Duplicates based on entity\_id were removed.

## 3. Uploading to MongoDB Atlas

Data from the final combined CSV file is uploaded to MongoDB Atlas using PyMongo. The connection uses the following credentials:

MongoDB Access URI (Read-Only):

mongodb+srv://guest:guest\_pass@intellewings.enj2zjz.mongodb.net/?retryWrites=true&w=majority&appName=Intellewings

📂 Access Code:

from pymongo import MongoClient  
  
# Connect to MongoDB Atlas (guest access)  
client = MongoClient("mongodb+srv://guest:guest\_pass@intellewings.enj2zjz.mongodb.net/?retryWrites=true&w=majority&appName=Intellewings")  
  
# Select the database and collection  
db = client["interpol\_data "]  
collection = db["red\_notices "]  
  
# Fetch and print all documents  
docs = collection.find()  
for doc in docs:  
 print(doc)