UNODC
United Nations Office on Drugs and Crime

UNODC Research

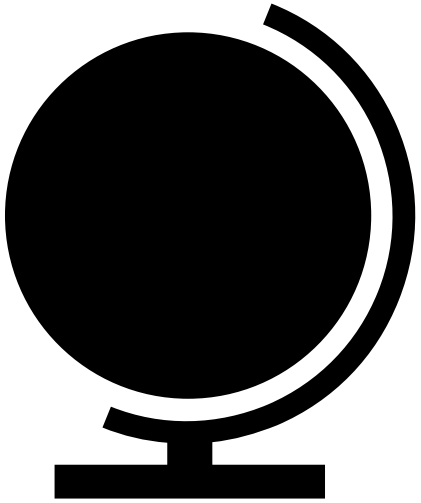
Scraping the web to improve drug seizure data

DATA DEVELOPMENT AND DISSEMINATION SECTION
RESEARCH AND TREND ANALYSIS BRANCH

UNITED NATIONS OFFICE ON DRUGS AND CRIME

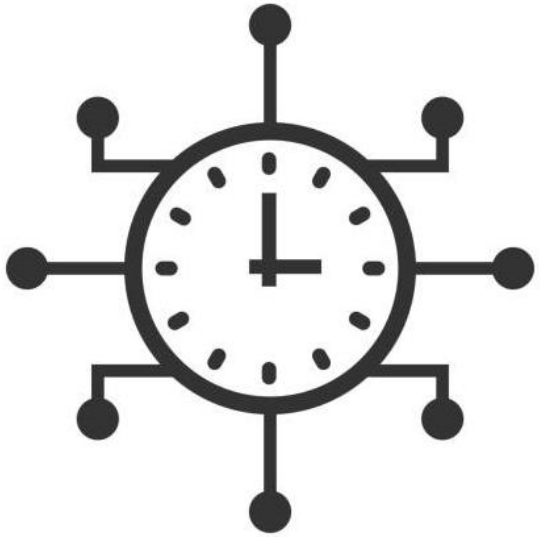
Drug Seizures: Primary Objectives for Real-time Data

Geographic Coverage



Linguistic Coverage

Real-Time Processing



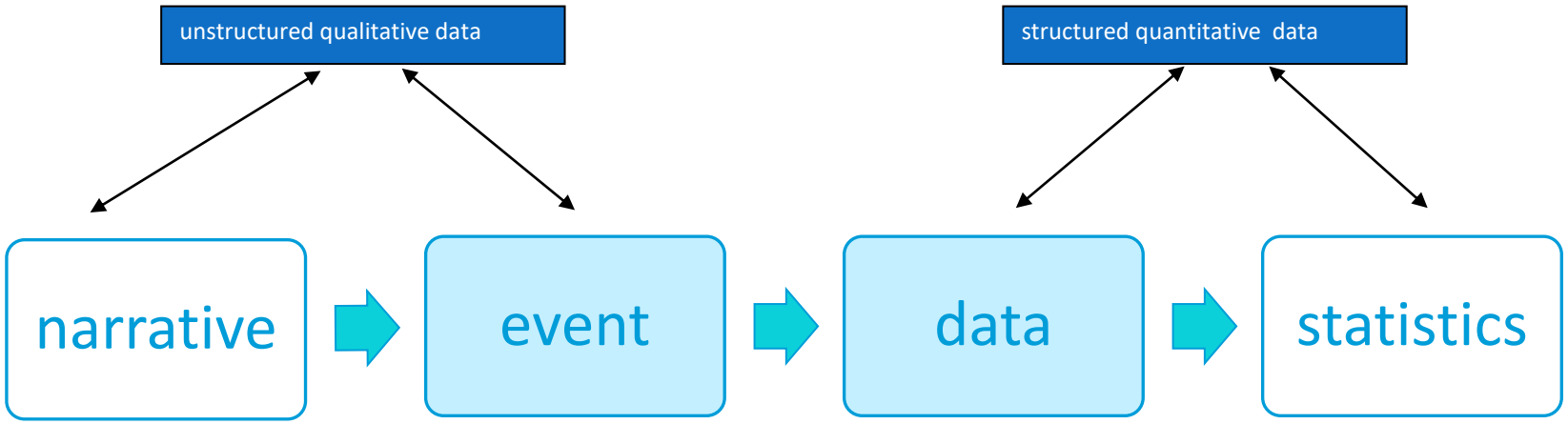
Quality Control

Complement existing data through Webscraping and Machine Learning Modeling techniques to transform unstructured data to improve overall timeliness and coverage of data and maintain quality

UNODC
United Nations Office on Drugs and Crime

UNODC Research

The challenge of Real-time/Open data processing



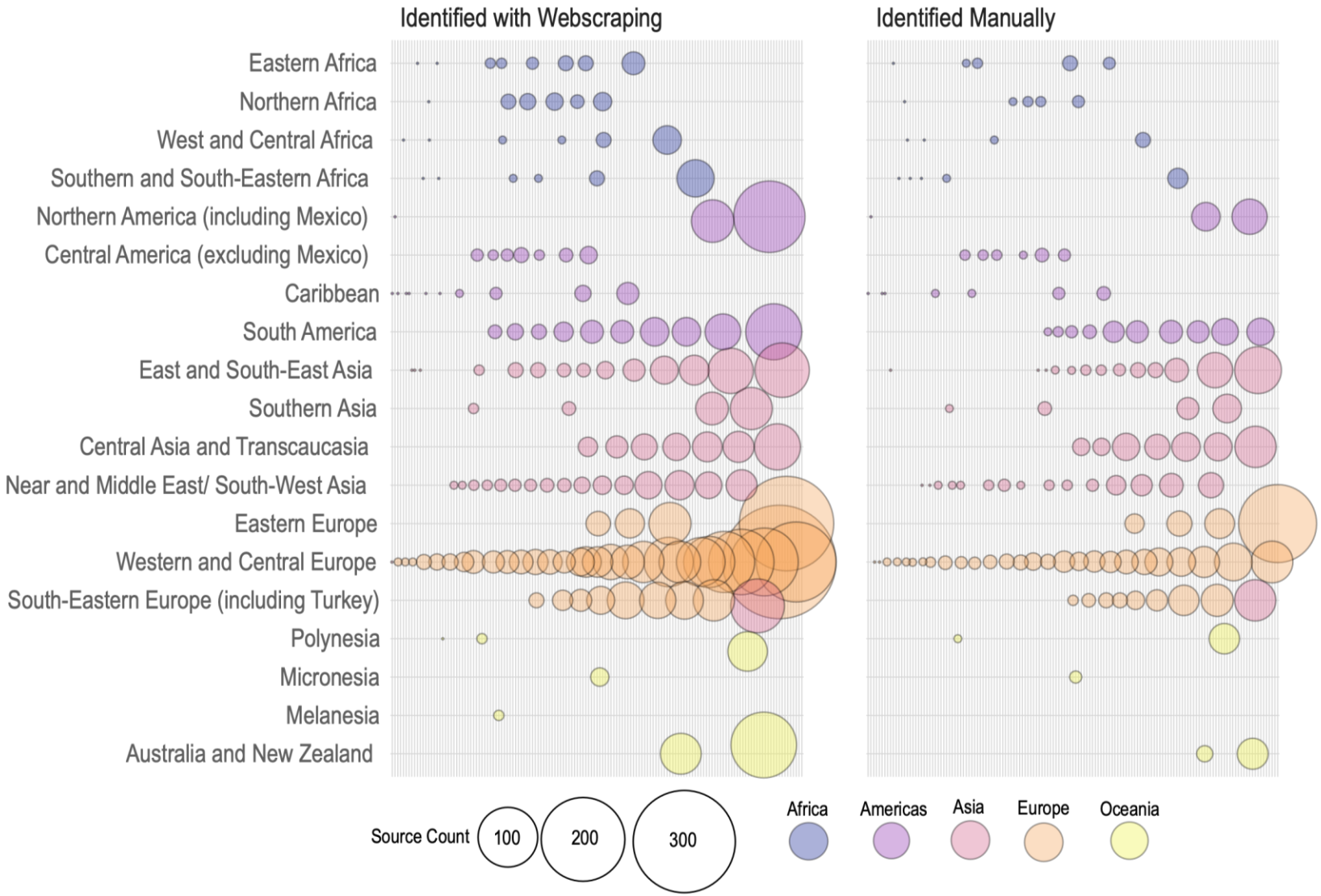
Sources: Geographic Coverage

Successes:

- Improved coverage in all regions
- 149 Countries
- Over 19,000 sources

Sources:

Publicly available public sector (Government) websites and news outlets, select social media

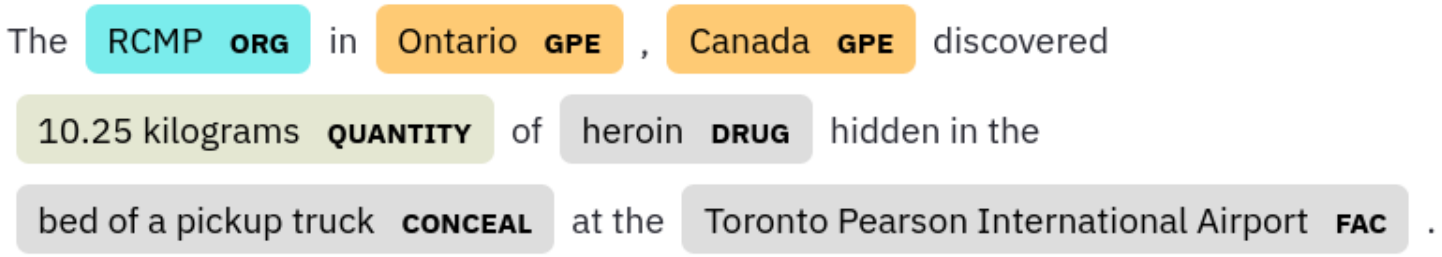


*One circle denotes one country. Size is determined by the number of unique news sources per country.

UNODC
United Nations Office on Drugs and Crime

UNODC Research

Real-time data processing: text prediction and extraction



- Possible to extract not only relevant features, but also features within a specific context.
- e.g. identify the authority responsible for drug seizure (Royal Canadian Mounted Police)
- Distinguish between a point of interest (Toronto Pearson International Airport) and the geopolitical entity (Ontario, Canada)
- Type of illicit drug, quantity, weight, location where the drug was concealed

UNODC
United Nations Office on Drugs and Crime

UNODC Research

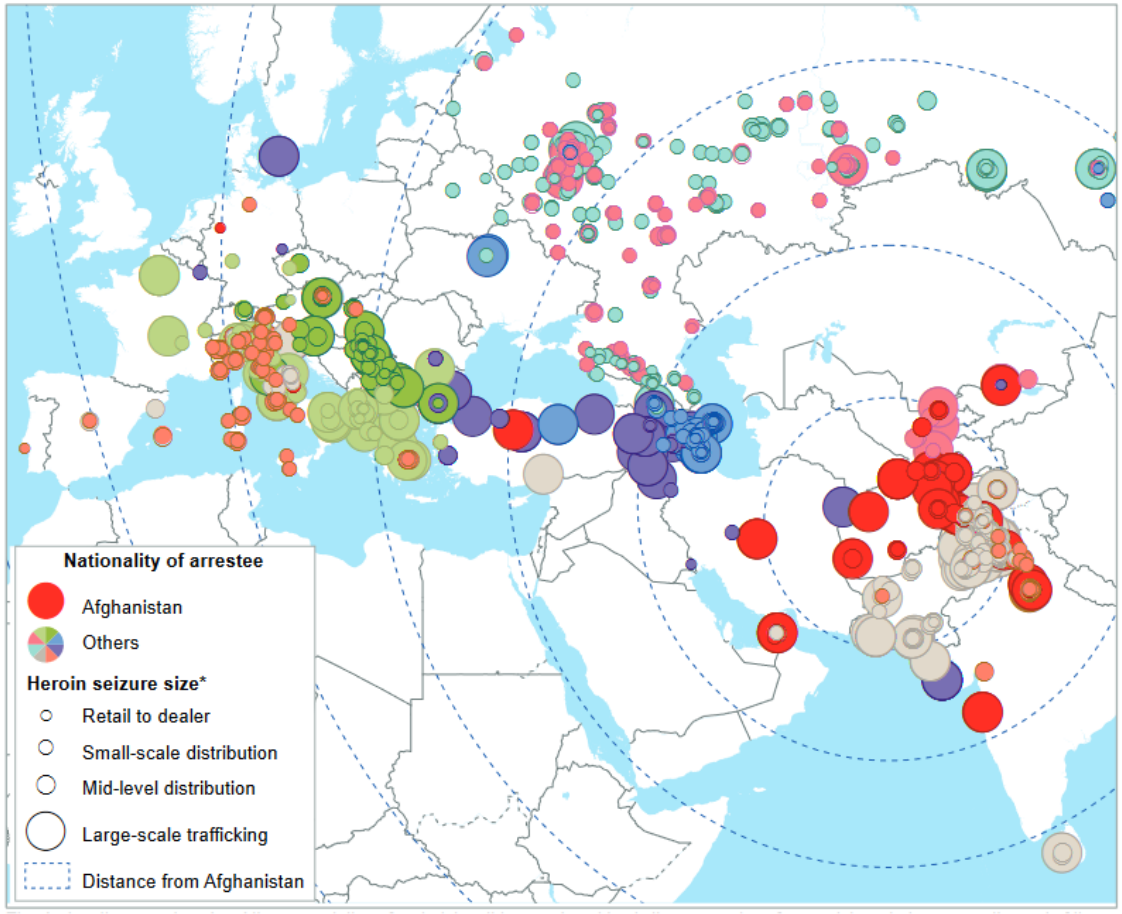
CHALLENGE

- Domain registration→ not all sites can be scraped
- Linguistic obstacles→ translation of over 53 languages from news articles
- Broader scope and volume of data→ processing requires updated machinery (AI/machine learning + IT infrastructure)
- Source quality

RESPONSE

- Improved data coverage in all regions
- Standardization of unit of analysis→ identification of common errors
- Annotation policy (control for relevance)
- 'Living libraries' feed the model + normalization (standards)
- Technique developed for eliminating copies of repeating data based on several event features

MAP 7: Nine most common traffickers by nationality involved in individual heroin seizures according to distance from Afghanistan and level of trafficking, January 2018–March 2021



Source: UNODC Drugs Monitoring Platform.

The designations employed and the presentation of material on this map do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted line represents approximately the Line of Control in Jammu and Kashmir agreed upon by India and Pakistan.

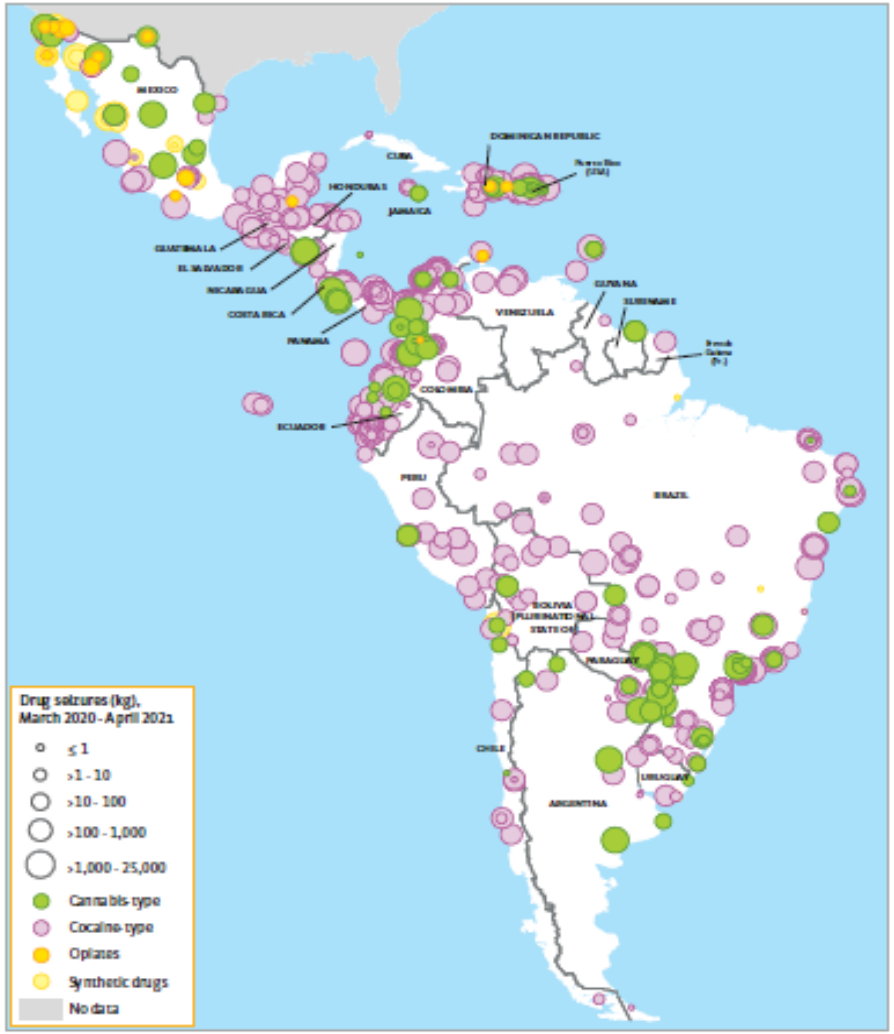
** Retail <0.05 kg // Small-scale 0.05-5 kg // Mid-level 5-50 kg // Large-scale > 50 kg.*

UNODC Research

Now open/real time data collection mainstreamed in our efforts to construct timely evidence: *Drugs Monitoring Platform (DMP)*

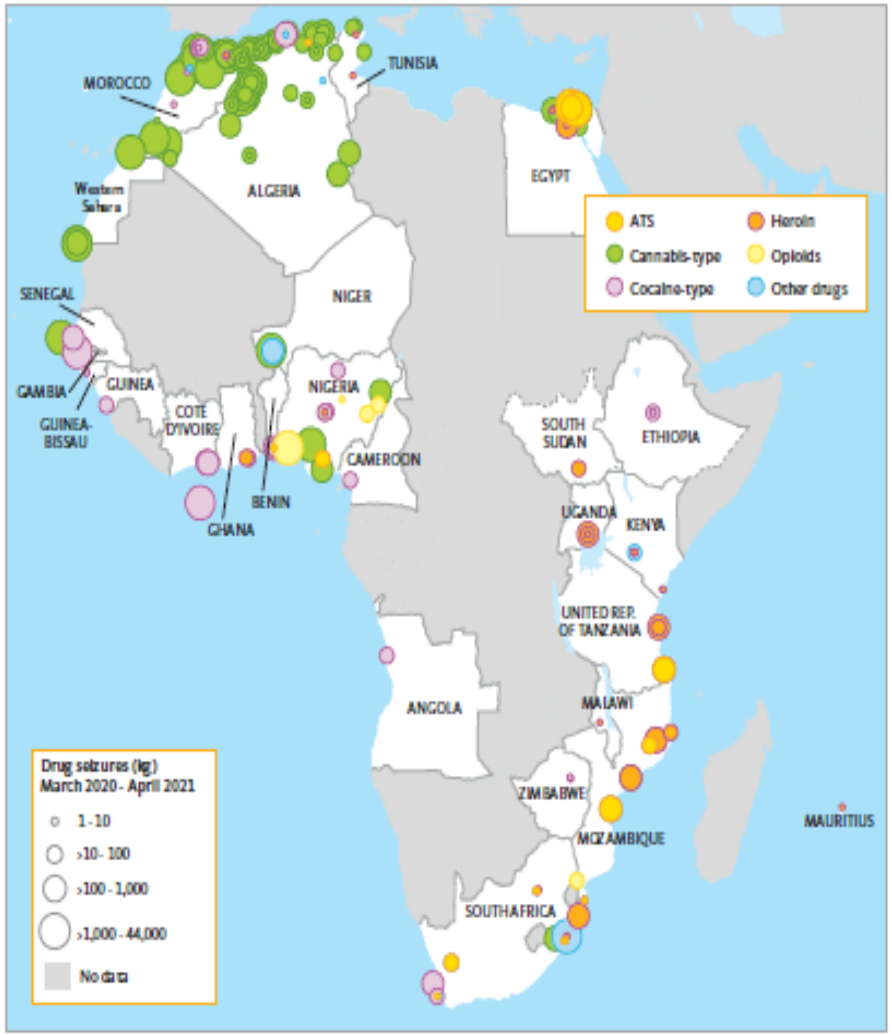
- Online geo-coded system with over 475,000 data points (secured data sharing)
- Collection, visualization and dissemination of data (interactive maps adapted to user needs)
- Monitoring of drug trafficking trends, routes, modus operandi, select trafficker data

MAP 4 Drug seizures in the Americas, March 2020–April 2021



Source: UNODC, Drugs Monitoring Platform.

MAP 9 Significant drug seizures in Africa, March 2020–April 2021



Source: UNODC, Drugs Monitoring Platform.

UNODC
United Nations Office on Drugs and Crime

UNODC Research

Thank you!

Further information available at:

- Drug Situation in Afghanistan (2021) https://www.unodc.org/documents/data-and-analysis/Afghanistan/Afghanistan_brief_Nov_2021.pdf
- UNODC Drugs Monitoring Platform Brief: The reach of nationally linked trafficking groups across Afghan opiate trafficking routes https://www.unodc.org/documents/data-and-analysis/Afghanistan/trafficking_groups_afghan_opiate_trafficking_routes.pdf
- UNODC Drugs Monitoring Platform Brief: Possible Impact of the COVID-19 pandemic on trafficking in opiates and methamphetamine originating in Afghanistan https://www.unodc.org/documents/data-and-analysis/covid/DMP_Brief_short.pdf

Francesca.massanello@un.org

Unodc-dmp@un.org