



Freedom
happens
now.



Considerations in Data Standardization for Operational Datasets

Sara Crowe
Data Analysis Program Director
April 8, 2019



The Human Trafficking Information Gap

OBSTACLES TO DATA COLLECTION

- Covert Crime and Barriers to Victim Identification
- Data Privacy Concerns
- Lack of Resources
 - Academic Studies
 - Operational Data

OBSTACLES TO DATA SHARING

- Data Privacy Concerns
- Incompatible Values and Variables
- Conflicting Definitions
- Incompatible Formatting
- Lack of Technical Resources and Solutions
- Bureaucratic Burdens on Data Collectors



Operational Data Benefits

- Utilizes existing systems and processes
- Continually updated
- Scalability
- May reflect broader geographic reach
- Provides insight into who is gaining assistance and efficacy of identification efforts
- Can inform the design of academic research studies

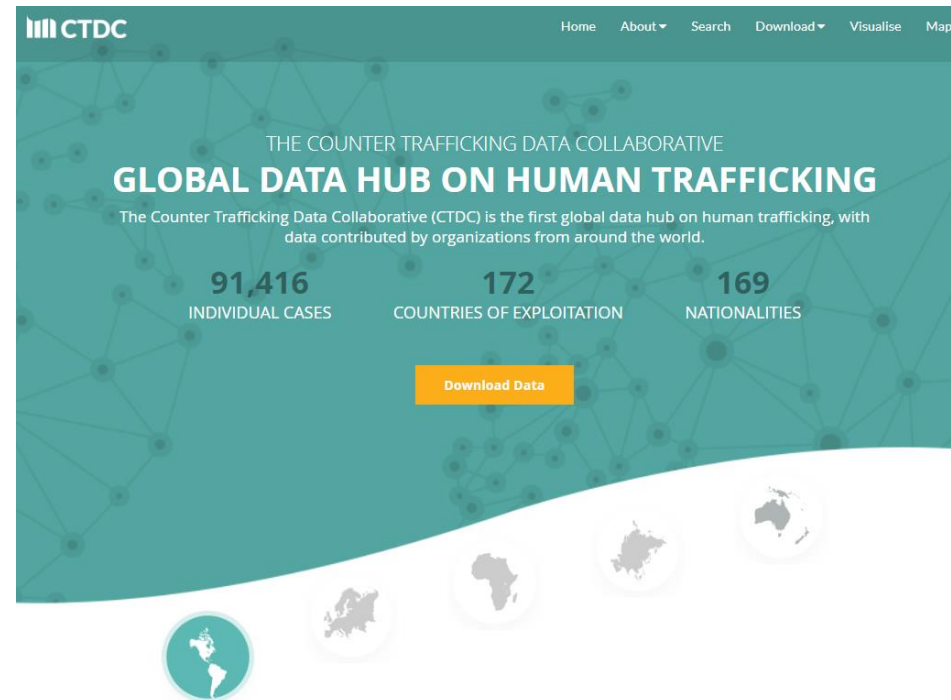


Operational Data Collection Challenges

- Reflective of awareness/access to support – Not prevalence
- Byproduct of another primary function
- Ethical issues surrounding information as a requirement to access assistance
- Strict definitions are not always relevant to response
- Data use and informed consent questions
- Processed insight vs. Info. attained during time of acute stress

CTDC Founding Ethos

- Privacy of data subjects is #1 priority
- Don't let the perfect be the enemy of the good
- Organizations with data need to drive the effort
- Increase in data contributors help include populations traditionally excluded
- Full transparency over limitations of data
- “Crowdsourcing” analysis yields greater impact
- “Big Data” mindset (if not yet reality)





Merging Datasets

POLARIS DATA

- PREVIOUSLY focused on “case records”

(Individual victim records since 1/1/2015)

- US legal definitions
- Reported through crisis hotline

(Assessments made but no standardized questions asked)

Victim Records
73 Shared
Variables

Aligned Values
and Definitions

IOM DATA

- Based on “victim records”
- International definitions
- Reported through client interviews



Initial Data Contributions



55,000
Victim
Records



24,000
Victim
Records



13,000
Victim
Records



Total Dataset: 91,000 records

- 169 Victim Nationalities
- 172 Countries of Exploitation



Two Primary Outputs

1. K-anonymized, disaggregate data for download

- Full codebook with information about data collection and standards available
- Data includes:

Demographics	Trafficking Context
Gender	Type of Trafficking
Age at Start of Exploitation	Industry of Exploitation
Age at Time of Contact with Agency	Means of Control
Citizenship	Country of Exploitation
Primary Language	Relationship to Recruiter

2. Interactive visual content using the full dataset, prior to k-anonymization

