Counter Trafficking Data Collaborative (CTDC)
Synthetic Data Codebook

Version 1 – September 2021

1.   DESCRIPTION OF DATA

These data consist of information on identified and reported victims of human trafficking. There are 37 variables that capture information on the socio-demographic profile of victims (such as gender or age), the trafficking process (such as means of control used on the victims), and the type of exploitation.

The Global Synthetic Dataset can be downloaded from the Counter Trafficking Data Collaborative (CTDC) website by researchers and statisticians who would like to undertake their own analysis. The Global Synthetic Dataset combines data from all CTDC partners (IOM, Polaris, Liberty Shared, A21, OTSH) and is put through two stages of de-identification. In the first stage, all names and identifying details are removed from the data. In the second stage, the dataset is processed through a privacy-preserving pipeline developed by Microsoft Research in collaboration with the International Organization for Migration (IOM).

This is the first synthetic dataset of victim of trafficking case records. Since July 2019, IOM has been working with Microsoft Research on this de-identification solution through the accelerator program of the technology industry coalition, Tech Against Trafficking (TAT). Prior to this, the CTDC team used k-anonymization to de-identify the dataset published on the website. As k-anonymization redacts cases with rare combinations of attributes marked as potentially identifying, the sample size can be reduced dramatically. In the case of our Global Dataset release, 40% of records were suppressed before release. The Microsoft algorithm overcomes the challenge of reduced sample size by synthesizing a new dataset in which records do not correspond to actual individuals, but which preserves the structure and statistics of the original data.

More information on the pipeline is available in a working paper and the open-source Python pipeline is available via GitHub. More information on k-anonymization is available here.

The Global Synthetic Dataset, made available for the first time in 2021, includes information on victims of trafficking. The synthesis algorithm ensures that estimated counts are equal to actual counts for individual attributes, and as close as possible for combinations of attributes. Privacy is preserved using a minimum reporting resolution k. The value of k is set to 10 following the way CTDC previously applied k-anonymity to the Global Dataset. In practice, this means that the synthetic data are generated to ensure that both privacy and utility are preserved. For instance, the number of victims trafficked for the purpose of forced labour will be the same in the raw and synthetic datasets. However, these aggregates will not allow "drill down" into groups of fewer than 10 real records, and they will be rounded to multiples of 10 to prevent inference over precise numbers (e.g., the last data release reported 100 victims

with certain attributes and the current release reported 101 victims, therefore a single victim with these attributes must have been added in the intervening period). The size of this dataset will increase as it will be regularly updated. The size of this dataset will increase as it will be regularly updated.

The size of the synthetic dataset differs from of the original dataset for two main reasons. First, attribute values are systematically redacted from case records such that the remaining combinations of attributes are common in the raw dataset and thus safe to release. As the raw dataset grows over time and with more partners, more and more of the case attributes should be able to safely filter through. Second, the algorithm creates new synthetic records from redacted attribute values to preserve the aggregate counts. It works by randomly selecting redacted attribute values in commonly-observed combinations. The random selection process continues to add other attributes until it would create a rare and thus potentially identifying combination. At this point, the synthesis process starts again for a new record. The process continues until no new records can be created without creating a rare combination of attribute values. As a final step, attribute values are randomly suppressed until the counts of attribute values in the synthetic dataset are a multiple of the privacy resolution k.

Please see Figure 1 below for a simple example.

Figure 1 – Creation of new synthetic records for $k=2$

Real data

gender	ageBroad	isForcedLabour
Female	18--20	1
Male	18--20	1
Male	18--20	1
Female	30--38	
Female	30--38	

Synthetic data

gender	ageBroad	isForcedLabour
Female		
	18 --20	1
Male	18--20	1
Male	18--20	1
Female	30--38	
Female	30--38	

The data included in the Global Synthetic Dataset is the result of an extensive process of comparing contributors' existing data classification systems, to identify data definitions which were identical or compatible. As CTDC co-founders, Polaris and IOM initially agreed on a shared lexicon and format. Where possible, CTDC follows international standard classification. Some concepts follow operational definitions as there are no international

agreed definitions, such as types of sexual exploitation.

Because of different data collection methods and different type of assistance offered by the contributing organizations to the Global Dataset, data are only available where the contributing organizations are operational and can share such data.

This dataset cannot be considered a random sample of all victims of human trafficking globally due to the hidden nature of this crime. Human trafficking cannot be easily recorded nor quantified. Potential victims of human trafficking need to be either identified or (self-) reported. This sample may be biased if certain types of trafficking or socio-demographic groups are more likely to be identified or referred than others. Since the unidentified population is unknown, the extent of bias is not known and cannot be corrected for.

IOM's data collection processes have evolved over time. Since 2011, identified and assisted victims of human trafficking are recorded in a web-based case management system. IOM records case data on victims of human trafficking in the context of its own programmes of direct assistance. The information of victims assisted by IOM has been entered into the Migrant Management Operational System Application (MiMOSA) - a web-based case management tool since 2011. MiMOSA has many data fields - from socio-demographic to route data - which serves the purpose of case management, reporting, and analysis. IOM contributes to the Global Synthetic Dataset through data from MiMOSA and "legacy" records from 2002 to 2011. More information on IOM's direct assistance activities and data is available [here](#).

The type of data collected reflects the organization's operations. IOM offers comprehensive direct assistance packages to victims of trafficking in collaboration with governments and civil society partners. The extent to which IOM has direct contact with victims and for which purposes vary. In some contexts, IOM identifies victims of trafficking and provides them with immediate assistance. In some countries, IOM effectively acts as the national referral mechanism for all identified victims of trafficking. In cases where victims have been trafficked out of their home countries, IOM could be involved in facilitating their safe, voluntary return to their countries of origin and ensuring long term assistance and re-integration, sometimes through referral to implementing partners in their countries of origin. In some cases, victims of trafficking would only be referred to IOM after they returned to their countries of origin. As the leading international organization providing direct assistance to victims of trafficking, IOM involves in bridging between different national contexts and facilitating the process of returning to home countries or resettlement to a third country.

Polaris' dataset is comprised of information obtained during Polaris' regular interactions with individuals contacting the U.S. National Human Trafficking Hotline and the BeFree Textline, both operated by Polaris. Polaris does not investigate claims made by individuals contacting the helplines and cannot verify the accuracy of the information reported. This data is not the result of a systematic survey. As these individuals told their own stories or relayed the experiences of their friends and family members, Polaris staff noted key elements of each account. This information was later classified in over 120 standardized fields using detailed

standards and definitions and this information constitutes the dataset contributed by Polaris. Victims and third parties reporting these situations were not asked a set of standardized questions and only provided information that they felt comfortable sharing with Polaris's staff to get the help they needed. As such, data contributed by Polaris only represents information disclosed to the National Human Trafficking Hotline and the BeFree Textline. Upon request, Polaris removes information about individuals who do not wish to be included in the dataset. More information about Polaris and its data is available [here](#).

Polaris has operated the National Human Trafficking Hotline since December 7, 2007 and the BeFree Textline since March 28, 2013. At present, Polaris is only able to contribute victim data from cases reported to the National Human Trafficking Hotline and the BeFree Textline since January 1, 2015 as the structure of data collected prior to this date is not compatible with CTDC's data standards. Polaris has reclassified its historical data and has been contributing data reported after March 31, 2017 on a biannual basis.

Liberty Shared's data collection process is centered around their Victim Case Management System (VCMS). The VCMS is a cloud-based data collection and information management tool, designed by Liberty Shared to assist frontline NGOs to combat human trafficking and modern slavery through robust record keeping. The VCMS brings together over 40 NGO partners working on this common challenge onto a shared platform. It facilitates standardized data collection. More information on Liberty Shared and its VCMS is available [here](#).

The VCMS has many data-fields available to assist frontline NGO users to record information related to victims and cases. Data points cover the critical components of a victim's experience, from pre-exploitation demographics, through the recruitment, transit and exploitation phases. NGO VCMS users receive training and ongoing support from Liberty Shared to record information.

The data contributed to the Global Synthetic Dataset by Liberty Shared comes directly from NGO partners utilizing the VCMS who have consented to be a part of CTDC. Some of this data has been directly entered by NGO users into the VCMS, which has been operational since 2014, whilst some data contributed by these partners is "legacy" data, migrated into the VCMS from their existing data storage setups when they transitioned onto the VCMS platform. Liberty Shared contributes data from our growing number of consenting partners on a bi-annual basis.

2. LIST OF VARIABLES

1 yearOfRegistration
2 gender
3 ageBroad
4 majorityStatusAtExploit
5 citizenship
6 CountryOfExploitation
7 traffickMonths
8 meansOfControlDebtBondage
9 meansOfControlTakesEarnings
10 meansOfControlThreats
11 meansOfControlPsychologicalAbuse
12 meansOfControlPhysicalAbuse
13 meansOfControlSexualAbuse
14 meansOfControlFalsePromises
15 meansOfControlPsychoactiveSubstances
16 meansOfControlRestrictsMovement
17 meansOfControlRestrictsMedicalCare
18 meansOfControlExcessiveWorkingHours
19 meansOfControlThreatOfLawEnforce
20 meansOfControlWithholdsNecessities
21 meansOfControlWithholdsDocuments
22 meansOfControlOther
23 isForcedLabour
24 isSexualExploit
25 isOtherExploit
26 typeOfLabourAgriculture
27 typeOfLabourConstruction
28 typeOfLabourDomesticWork
29 typeOfLabourHospitality
30 typeOfLabourOther
31 typeOfSexProstitution
32 typeOfSexPornography
33 typeOfSexOther
34 recruiterRelationIntimatePartner
35 recruiterRelationFriend
36 recruiterRelationFamily
37 recruiterRelationOther

3.   DETAILED DESCRIPTION OF VARIABLES

Before undertaking exploratory analysis, please consider that a significant number of missing values for a variable may be a result of the fact that only certain CTDC contributors have data collected.

Variable 1
Variable label: yearOfRegistration
Type: numeric
Values and categories:
- Range: [2002, 2021]

Definition: The year in which the individual was registered and assisted by IOM or the year in which Polaris first received a report of the individual's trafficking situation, or the date of entry into Liberty Shared's VCMS.

Variable 2
Variable label: gender
Type: string
Values and categories:
- NULL *[missing data]*
- Male [*Individual's sex assigned at birth is male and the individual identifies as male*]
- Female [*Individual's sex assigned at birth is female and the individual identifies as female*]
- Transgender/NonConforming [*The individuals's expression of gender falls outside binary societal expectations of what a man and a woman are. Including, but not limited to: Individuals whose sex assigned at birth is female, but the individual identifies as male; Individuals whose assigned at birth is male, but the individual identifies as female; Individuals who identify as both male and female or neither male nor female. For data privacy reasons, these data are not further disaggregated in this dataset.*]

IOM started collecting data about gender (and not sex) which includes transgender categories only in 2017.
Definition: Designates the individual's expression or condition of being male, female, both or neither. This category considers the psychological, behavioral, social, and cultural aspects of being male, female, or non-binary (masculinity, femininity, neither, or a combination of both).

Variable 3
Variable label: ageBroad
Type: string
Values and categories:
- NULL *[missing data]*
- 0—8
- 9—17
- 18—20
- 21—23
- 24—26
- 27—29
- 30—38
- 39—47
- 48+

Definition: The individual's age at the time the individual was registered and assisted by IOM or Polaris first received a report of the individual's trafficking situation or at first point of contact with Liberty Shared's NGO partner.

Variable 4
Variable label: majorityStatusAtExploit
Type: string
Values and categories:
- NULL *[missing data]*
- Minor   [*Any person under the age of 18*]
- Adult   [*Any person who is 18 or older*]

Definition: The individual's majority status when exploitation first began. Exploitation includes the exploitation of the prostitution of others or other forms of sexual exploitation, forced labour or services, slavery or practices similar to slavery, servitude or the removal of organs.

Variable 5
Variable label: citizenship
Type: string
Values and categories:
- NULL *[missing data]*
- Values based on ISO 3166-1 Alpha-3 Codes

Definition: The set of rights and duties that a person has with a country because of his/her legal bond with the country. This term is often used interchangeably with nationality. Liberty Shared records this as 'Country of Origin' as a proxy for citizenship. For cases with more than one citizenship, this dataset publishes the unique citizenship value based on the first reported field. Please refer to **Appendix** for more information**.**

Variable 6
Variable label: CountryOfExploition
Type: string
Values and categories:
- NULL *[missing data]*
- Values based on ISO 3166-1 Alpha-3 Codes

Definition: Indicates the country where a victim is first supported/assisted, identified and/or referred. In the context of human trafficking data, this is also referred to as the "country of destination2 (as opposed to "country of origin" if human trafficking was across borders). In the case of IOM, the last country of exploitation has a high number of missing values. Therefore, a proxy has been created to capture as much data as possible about the country the victim was exploited in. Please refer to **Appendix** for more information.

Variable 7
Variable label: traffickMonths
Type: string
Values and categories:
- NULL *[missing data]*
- 0—12 (0-1 yr)
- 13—24 (1-2 yrs)
- 25—60 (2-5 yrs)
- 61+ (5+ yrs)

Definition: Indicates the reported duration of trafficking in months.

Variable 8
Variable label: meansOfControlDebtBondage
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual is forced to work to pay off a created or perceived debt. The individual is deceived to work for little or no pay, with no control over his/her debt. Debt bondage is defined as the status or condition arising from a pledge by a debtor of his personal services or of those of a person under his control as security for a debt, if the value of those services as reasonably assessed is not applied towards the liquidation of the debt or the length and nature of those services are not respectively limited and defined (United Nations' 1956 Supplementary Convention on the Abolition of Slavery).

Variable 9
Variable label: meansOfControlTakesEarnings
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual has experienced a situation where the exploiters have taken his/her remuneration in order to control him/her.

Variable 10
Variable label: meansOfControlThreats
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual experienced a situation in which his/her exploiter(s) explicitly or implicitly communicated an intent to inflict harm or loss on the individual or another. Threats to report or contact law enforcement in order to negatively impact the individual or another are considered distinct and are categorized in the *meansOfControlThreatOfLawEnforcement* variable.

Variable 11
Variable label: meansOfControlPsychologicalAbuse
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual experienced a situation in which their exploiter(s) used emotionally abusive, deceptive, or devious tactics to influence the individual. This may include, but is not limited to, name calling, verbal abuse, humiliating in front of others, manipulating real or perceived power imbalances, or shaming, blaming the individual. It may also include acts intended to exploit or prey upon any familial or romantic bonds/attachments the individual has with their exploiter(s).

Variable 12
bel: meansOfControlPhysicalAbuse
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual experienced a situation in which their exploiter(s) acted to cause physical injury, pain, disability, death or trauma to the individual. This includes but not limited to shoving, strangulation, shaking, slapping, punching, kicking, pulling hair, burning, branding or tattooing, the use of a weapon, or using one's size and strength against the individual. Physical acts of a sexual abuse are considered distinct and are categorized in the *meansOfControlSexualAbuse* variable.

Variable 13
Variable label: meansOfControlSexualAbuse
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual experienced any kind of unwanted or non-consenting sexual contact from their exploiter(s) as a means by which to control the individual, rather than a purpose for which the individual was trafficked, in order to influence their behavior. This includes, but is not limited to, using sexual acts, assault, or contact as punishment or manipulation or normalizing sexual violence. It also includes coercive behavior that interferes with the individual's ability to control his/her reproductive life, including but not limited to, forcing/coercing the individual to terminate or continue a pregnancy against their will, manipulating birth control, intentionally exposing someone to an STI, preventing condom negotiation, and/or attempting to or impregnating the individual without their consent.

Variable 14
Variable label: meansOfControlFalsePromises
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual was defrauded or tricked into entering the exploitative situation by their exploiter(s) using deception and false pretenses.

Variable 15
Variable label: meansOfControlPsychoactiveSubstances
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the exploiter(s) induced the individual into substance abuse, provided substances to make the individual compliant or in order to influence their behavior, or exploited an existing substance abuse issue.

Variable 16
Variable label: meansOfControlRestrictsMovement
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the exploiter(s) isolated, confined or limited the movement of the individual in any way physically or socially. This may include situations in which the individual is physically detained, prevented from moving without being accompanied or monitored, or the exploiter(s) threatens or enacts repercussions related to the individual's movement. This may also include forms of emotional isolation including restricting the individual's access to support systems or social networks or moving the individual frequently to prevent the individual from establishing support systems or social networks.

Variable 17
Variable label: meansOfControlRestrictsMedicalCare
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the exploiter(s) limited the individual's access to medical or health services. Including but not limited to necessary medical care or treatment being withheld, or when access to such treatment was controlled by the exploiter(s). This category also includes situations in which the individual was unable to access or interact with health services without being accompanied or monitored by the exploiter(s).

Variable 18
Variable label: meansOfControlExcessiveWorkingHours
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual was required to work a significant number of hours more than what they were contracted or promised; could include over-time, late or atypical shifts, or overnight hours. Could be used a means for keeping the individual isolated and/or unable to seek help or report their situation. In some instances, work hours may be so excessive as to cause physical and/or mental health issues which may also limit the individual's capacity to seek help or report their situation.

Variable 19
Variable label: meansOfControlThreatOfLawEnforcement
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual experienced a situation in which their exploiter(s) explicitly or implicitly communicated an intent to contact or involve law enforcement or other relevant authorities, such as immigration authorities, in order to negatively impact the individual or another.

Variable 20
Variable label: meansOfControlWithholdsNecessities
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual experienced a situation in which their exploiter(s) denied, restricted, or threatened to deny/restrict basic living necessities such as food, shelter, water, hygiene, appropriate clothes, or necessary items for religious observance or gender expression.

Variable 21
Variable label: meansOfControlWithholdsDocuments
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual experienced a situation in which their exploiter(s) limited, restricted, or controlled the individual's access to important documents including, but not limited to, the individual's passport, immigration documents, work permit, identification card, government benefit documents, birth certificate, gender identity carry letter, court-issued protection orders, custody papers, or other legal, official, or government documents.

Variable 22
Label: meansOfControlOther
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the exploiter used a tactic intended to create or maintain power and control over the individual that cannot reasonably fit in any of the previous categories in **Variables 8-21**. Two variables in the CTDC k-anonymized dataset, "meansOfControlRestrictsFinancialAccess" and "meansOfControlUsesChildren", are grouped as other due to low response rates.

Variable 23
Variable label: isForcedLabour
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: It indicates that the purpose for which a victim was trafficked was all work or service which was exacted from the individual, under the threat of a penalty and for which the individual has not offered himself or herself voluntarily. Sexual services are excluded from this definition.

Variable 24
Variable label: isSexualExploit
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: It indicates that the purpose for which a victim was trafficked was sexual services, such as the exploitation of the prostitution of an individual. Typically, the exploiter(s) use force, fraud or coercion to achieve exploitation.

Variable 25
Variable label: isOtherExploit
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: It indicates that the purpose for which a victim was trafficked was other type of exploitation, including those who underwent forced marriage, forced military, and organ removal. Typically, the exploiter(s) use force, fraud, or coercion to achieve exploitation. Polaris staff apply the U.S. federal definition of human trafficking as defined by the Trafficking Victims Protection Action (TVPA) to determine if a situation described through the helplines has indications of human trafficking. Situations of forced marriage, organ harvesting or "other", which do not meet the U.S. definition of trafficking are not included in the data contributed by Polaris to the CTDC as Polaris currently does not collect data about these sub-types.

Variable 26
Variable label: typeOfLabourAgriculture
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual experienced forced labour in activities defined in ISIC Section A Division 01, "Crop and animal production, hunting and related service activities". This category does not include work related to forestry and logging as defined by ISIC Section A, Division 02 which is considered distinct and categorized in **Variable 30** *typeOfLabourOther*.

Variable 27
Variable label: typeOfLabourConstruction
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual experienced forced labour in activities defined in ISIC Section F "Construction".

Variable 28
Variable label: typeOfLabourDomesticWork
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual experienced forced labour in activities defined in ISIC Section T "Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use".

Variable 29
Variable label: typeOfLabourHospitality
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual experienced forced labour in activities defined in ISIC Section I "Accommodation and food service activities" including both Division 55 "Accommodation" and Division 56 "Food and beverage service activities".

Variable 30
Variable label: typeOfLabourOther
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual experienced forced labour of a specified type which could not reasonably be classified in any of the categories specified in **Variables 26-29**. Due to low response rates, the following variables in the CTDC k-anonymized data are grouped as other: aquafarming, begging, illicit activities, manufacturing, mining or drilling, peddling, and transportation and storage.

Variable 31
Variable label: typeOfSexProstitution
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual experienced exploitation in activities associated with an individual's sexual act for payment.

Variable 32
Variable label: typeOfSexPornography
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual experienced exploitation in the production of visual material depicting sexual behavior that is intended to arouse sexual excitement in its audience and does not involve any participation from the audience. It does not include situations in which the audience remotely participates or interacts with the individual featured in the visual material.

Variable 33
Variable label: typeOfSexOther
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether the individual experienced exploitation depicting sexual behavior other than prostitution and pornography. It includes remote interactive services and private sexual services in the CTDC k-anonymized data.

Variable 34
Variable label: recruiterRelationIntimatePartner
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether a person who initially enticed or obtained the individual into the situation of exploitation was one with whom the individual has identified having a current or former romantic relationship.

Variable 35
Variable label: recruiterRelationFriend
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether a person who initially enticed or obtained the individual into the situation of exploitation was one with whom the individual was familiar, exclusive of romantic partners, family relations, or other more formal relationships.

Variable 36
Variable label: recruiterRelationFamily
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether a person who initially enticed or obtained the individual into the situation of exploitation was one with whom the individual was connected biologically, through marriage, or a person who the individual has identified as having been their current or former custodian or guardian. This may include, but is not limited to, parents, primary caregivers, and foster parents.

Variable 37
Variable label: recruiterRelationOther
Type: binary numeric
Values and categories:
- NULL *[missing data]*
- 1

Definition: Indicates whether a person who initially enticed or obtained the individual into the situation of exploitation was a person with whom the individual had any other notable relationship that cannot reasonably fit into previous categories. This may include, but is not limited to, labor brokers, contractors, formal employers, or smugglers. In the case of Liberty Shared, data collected on recruiters was more extensive than Global Dataset required. Liberty Shared transformed values to match IOM values, resulting in a higher level of 'other' values where information recorded did not match (for example – Broker – becomes other).

Appendix

- General comment on variables 5 (citizenship) and 6 (CountryOfExploitation)

  - Serbia: includes Kosovo following UN guidelines.
  - China: includes Tibet.

- Citizenship (Variable 5 – citizenship)

  Multiple citizenships are recorded for victims from Polaris. Victims identified/referred/assisted by IOM have only one main citizenship recorded. In this dataset, we use the unique value that is the first indicated citizenship.

- Country of exploitation proxy variable (Variable 6 – CountryOfExploitation)

  This is usually the country where a victim is first identified and/or referred to IOM. This does not necessarily capture the last part of the victim's journey before they are identified and referred nor the intended country of destination, which is not currently captured in the IOM database. The data points that are recorded in the case management system are 'last country of exploitation', 'location of the screening interview', 'location of the victim upon registration', and 'location of the IOM mission registering the case'.