



The Central Asia Political Exiles

Codebook

University of Exeter

Project Team

Saipira Furstenberg, University of Exeter

John Heathershaw, University of Exeter

Edward Lemon, Daniel Morgan Graduate School

David Lewis, University of Exeter

Alexander Cooley, Columbia University

Contents

1. Introduction and Overview	4
1.1 Source of Funding	4
1.2 Partner organisations	4
1.3 Methods	4
1.4 Context dependence of the data	5
1.5 Empirical Challenges of the Research Topic	5
1.6 Training	6
2. Definitions	7
2.1 Who is a political exile?	7
2.2 What are ‘repression, ‘transnational repression’ and ‘extraterritorial security’?	8
2.3 What is physical violence?	9
2.4 What is state of concern?	9
2.5 What is gender?	9
2.7 What is year?	9
2.8 What is alleged affiliation?	10
2.9 What are countries of transit, asylum and/or residence?	10
3. Coding and Methods	10
3.1 Variables and Codes	10
3.1.1 State of concern (Nominal)	10
3.1.2 Year	10
3.1.3 Types of political exile (Nominal)	10
3.1.4 Stages of transnational repression (Ordinal)	11
3.1.5 Gender (Nominal)	11
3.1.6 International arrest warrants (Nominal)	11
3.1.7 Physical Violence (Ordinal)	11
3.1.8 Further notes	12
3.2 Sources and Methods	12
3.3 Ethics	12
4. Production	13
4.1 Production and Analysis of the Dataset	13
4.1.1 Add new entries based on searching sources – see 3.2	13
4.1.2 Identify categories for each new entry’s fields – see 3.1	13
4.1.3 Identify incident stages for each new entry’s fields – see 3.1	13

5.Verification.....	13
6.Update.....	14
6.1 Supervisor check.....	14
7. Review Process.....	14
7.1 Peer review to check, supplement and validate entries.....	14
7.2 Revisions are made following review.....	14
7.3 Consider new entries recommended by reviewers.....	14
7.4 Internal review.....	14
7.5 Analysis and commentary.....	15
7.6 A new edition is released for publication.....	15

1. Introduction and Overview

This document describes the Central Asia Political Exiles (CAPE) dataset, a data collection project of the Central Asian Studies network at the University of Exeter (ExCAS). It was initiated in October 2014 by John Heathershaw and Alexander Cooley in partnership with David Lewis and Ed Lemon. Saipira Furstenberg joined the project team in 2017 as project manager. The dataset offers a unique analytical tool to study the dynamics of extraterritorial security measures (sometimes called ‘transnational repression’) in countries of Central Asia to target dissidents abroad.

The CAPE dataset describes and categorises transnational repression carried by the five states of Central Asia from 1990-2018 on five categories of political exile and across three stages of intensity. The dataset is available for download at: <https://excas.net/exiles/>

The database serves as a place of the collation and analysis of publically-available data – a one-stop shop to learn about political exiles and, more importantly, the patterns of extra-territorial security to which they are subject. The database is reviewed and updated on a regular basis according to principles of method and ethics. The updates include some changes in the data published for previous years, consequently the database is adjusted accordingly.

The collection of the material in the database is extracted from events being reported in news media and official policy and judicial reports published either in Russian and/or English languages. In other words, the database only includes information already made public.

1.1 Source of Funding

Work on the dataset has been supported by a grant of the Open Society Soros Foundation and internal funding from the University of Exeter.

1.2 Partner organisations

We cooperate with the following organizations in data sharing and/or joint advocacy initiatives:

- Amnesty International
- Fair Trials International
- Foreign Policy Centre
- Freedom House
- The Human Rights Center “Memorial”
- Civic Assistance Committee
- Human Rights Watch
- Associate for Human Rights in Central Asia (AHRCA)

1.3 Methods

The large size of the database meant that for us to code the data within the usual time restrictions of the granting process, we were going to need a large staff working to enter the data. The budget restrictions meant that we were going to be severely limited, we decided to solve this issue by employing undergraduate volunteers and interns. We took the decision, to develop a web-based data entry system that would allow a very large number of students to work on the database, using their own equipment, on a flexible schedule. This method also has the advantage of giving us a good deal of control over the data entry process: the computerized record gives us information of how much time all of our data coders were putting in. Additionally, the system allows us to verify individual coding records for accuracy.

1.4 Context dependence of the data

When dealing with a data collection that spans several years, it is important to be aware that political changes or events can have a notable impact on observable year-to-year trends in the data. The CAPE database covers the period of 1990-2018 inclusive. The variability of recorded cases is highly dependent on the political context of the countries of origin.

For example, the CAPE team noticed a sizable increase in the frequency of political exiles incidents in year 2015 for Tajikistan, and this trend continued through to the end of the 2017. The increase of incidents in Tajikistan corresponds to the forced closure of the Islamic Renaissance Party of Tajikistan (opposition party) in the country in 2015. On September 29, 2015, at the request of the Tajik Prosecutor General, the High Court classified the IRPT as a terrorist group. As a result of the political turmoil, party members went into exile fearing for their lives.

In similar vein, we have recorded high level of incidents in 2005 for Uzbekistan which correspond to the Andijan event and 2010 for Kyrgyzstan, a year marked by ethnic clashes in Osh. The amount of data collected for a specific country also highly dependent on the available documentation. These include media articles, online policy reports and secondary source materials such as legal documents. There is a high volume of data collected on political exiles for countries of Tajikistan and Uzbekistan as their public political exile movements are larger and, especially in Tajikistan's case, have emerged in the age of the internet.

Several efforts were also made to collect public information from before the internet era in cooperation with civil society activists who have followed these cases since the 1990s. We worked closely with the Human Rights Centre "Memorial" and the Civic Committee Assistance in Moscow to collect additional exiles cases. The cases were then searched online in order to record them in the database. There is little data available on political exiles in Turkmenistan due to the limited published material sources on this topic.

1.5 Empirical Challenges of the Research Topic

Beyond the specific Central Asian contextual challenges, research on the topic of political exiles is difficult for many reasons. Perhaps the most challenging factor is that our targeted population constitute so-called 'hidden populations'. According to Tyldum and Brunovskis¹ (2005, p. 18), a 'hidden population', 'is a group of individuals for whom the size and boundaries are unknown, and for whom no sampling frame exists'. Additionally, conducting research on 'closed' and 'authoritarian' contexts, respectively represent itself a challenge (Koch, 2013²; Glasius et al, 2018³). Tensions between transparency obligations and protection of respondents are particularly acute when it comes to fieldwork research in authoritarian circumstances (Glasius et al, 2018, p.3). For empirical studies, such challenges require approaches different from those commonly used for more easily observable populations. The political and coercive environment of authoritarian states demands a great vigilance to data collection and source interpretation to avoid bias and politicization of research (Bekmurzaev, Lottholz, & Meyer, 2018⁴).

¹ Tyldum, G. and Brunovskis, A. (2005). 'Describing the Unobserved: Methodological Challenges in Empirical Studies on Human Trafficking'. *International Migration*, 43 (1/2), pp. 17-34.

² Koch, N. (2013a). Introduction—Field methods in "closed contexts": Undertaking research in authoritarian states and places. *Area*, 45, 390–395.

³ Glasius, M. et al. (2018). *Research, Ethics and Risk in the Authoritarian Field*. Palgrave.

⁴ Bekmurzaev, N., Lottholz, P. H., & Meyer, J. (2018). Navigating the safety implications of doing research and being researched in Kyrgyzstan: Cooperation, networks and framing. *Central Asian Survey*, 37, 100–118

Our database is constituted of online published material. It's important to note that we only gather information that is public and is identifiable, but not private. The use of online source material also allows us to reach out and conduct follow-up studies within the populations that may be hard to reach in traditional research.

Due to the nature of collected data, the dataset is likely to have a substantial effect on the results of empirical research. Our reliance on open sources means that, while representative of the public record, the database is inherently limited and can make no necessary claim to be comprehensive. Academics and analysts using the database should therefore consider that, as a tip of the iceberg, it may not be representative of the whole.

Our data was also collected at different points of time by different research assistants. This can potentially create empirical discrepancies. To limit discrepancies, we have created definitions on specific variables to be collected and have established coding guidelines to help members to identify relevant data. This approach enabled the research assistants to perform targeted searches and filter the relevant information on political exiles. Team members ensured the quality of coded data was accurate by reviewing systematically the work of research assistants.

To establish the reliability of the data collection we have used the triangulation method that involved the use of multiple independent sources of data to establish the truth and accuracy of the information. This process involved checking the information against a well-known targeted source both in English and Russian languages. Additionally, in some cases we would consult our expert group composed of civil society members working on the topic of political exiles from Central Asia.

Our dataset only includes the period 1990-2018. The time span, beginning in the year that Central Asian states declared their sovereignty within the USSR, is a data base of post-Soviet "transition". In terms of country coverage, the data only deals with the five Central Asian countries it therefore provides a small regional case study of political exiles, extraterritorial security, and "transnational repression".

Variation in incidents levels and number of recorded exiles cases can also be explained by the relationships that allowed us to identify lesser-known public records. For instance, we have identified historical data for Uzbekistan and Turkmenistan via our civil society experts for the early 90s; however, little information was provided for Tajikistan, Kyrgyzstan and Kazakhstan for that timeframe.

1.6 Training

Volunteer student Research Assistants were recruited via the University of Exeter MA Conflict, Security and Development research assistantship programme. They were invited to attend a training course where the project manager and supervisor explained the nature of the original CAPE data and how the data had been collected, explained the goals of the current project related especially to data entry, offered detailed explanations of the data-entry codebook including examples of data entry, and discussed administrative and ethical procedures for working on the project. Students were given time to practice data entry with the project supervisor. Each student was then asked to enter the same 20 test cases over within the following week. The project staff stressed to the students that all data entry mistakes should be identified by students without fear of penalty, and that any unusual or confusing data encountered should be brought to the attention of supervisory project staff.

2. Definitions

The CAPE aims to measure patterns of transnational repression carried by the states of Central Asia (Uzbekistan, Kazakhstan, Kirgizstan, Turkmenistan, Tajikistan) on their population abroad. The database contains 5 categories of exile (see categories of exile definition 2.1) and 3 stages of incident types (see definition extraterritorial security 2.2). Throughout the years the language and the data entry rules became increasingly detailed and specific. For example, we have created specific rules for using the value “unknown” when there was no information provided on the given value or field. Additionally, we have added additional entries to characterise the types of violence, arrests notice and gender characteristics. These revisions and additions to the codebook and web- interface were all made in the interest of increasing data entry reliability while decreasing data entry error.

2.1 Who is a political exile?

We define a political exile as an emigrant who has settled or spent a prolonged period overseas for reasons which are wholly or partly of a political character.

This category of individuals in political exile bears one or more of the following characteristics:

- Individuals who have previously acted politically, in government or in opposition, in a role which has pitted them against the regime of power, and/or individuals who have previously opposed their government from outside of formal politics.
- Individuals of civil society, facing the danger of persecution in their home based country for political reasons.
- Individuals participating in exile politics — attention here is on exile’s political activities while abroad and its implication on regime’s domestic politics.
- Individuals persecuted for their social affiliation or the political activities of their relatives or friends.

In the CAPE database we identify five categories of political exile which are observable in a Central Asian context:

1. Former regime insiders and their family members.
2. Members of opposition political parties and movements.
3. Banned clerics and alleged religious extremists.
4. Independent journalists, academics and civil society activists.
5. Others: businessmen, employees or relatives of political exiles.

Excluded from our database are the following three categories of person:

1. War criminals and individuals convicted of terrorist offences overseas in a court in a jurisdiction where a high standard of the rule of law is upheld. None of the current legal processes with regard to terrorism offences within any of the five Central Asian republics can be considered as upholding a high standard of the rule of law.
2. Members of transnational clandestine groups, including proscribed terrorist organizations. This includes any individual where there is a credible public source (i.e. established media in

an environment with freedom of press) reporting them as a member of a proscribed group (according to the US list or another plausible list).

3. Labour migrants. Migrants subject to bureaucratic control and even 'administrative persecution' do not count as political exiles unless there is a prima facie political rationale for these measures taken against them (i.e. they are members of a group which attributes them to one of the five categories of political exile).

2.2 What are 'repression, 'transnational repression' and 'extraterritorial security'?

Christian Davenport defines repression as "actual or threatened use of physical sanctions against an individual or organization, within the territorial jurisdiction of the state" (Davenport 2007: 2).

Transnational repression arises when governments target their opponents in territories and spaces beyond their home country (Moss 2016). These governments seek control not just of their own territories but any spaces, both physical and virtual, where their political opponents and co-ethnic diaspora are found. Their targets are principally their own domestic political opponents.

'Extraterritorial security' is our own concept of what is elsewhere denoted as 'transnational repression'. For authoritarian governments, extraterritorial security practices form an extension of their domestic pursuit of regime security. In other words, extraterritorial security involves "the practice of internal security within the territory of a foreign state" (Cooley and Heathershaw 2017: 191). The CAPE database identifies three stages of extra-territorial security developed inductively from the study of Central Asian political exiles. Each denotes an incident which indicates an escalation of action taken against the exile.

Stage 1

Put on notice, which includes informal warnings and threats to individuals and intimidation of family members and formal arrests warrants, including Interpol notices, and extradition requests

Stage 2

Arrest and/or detention, which includes short-term and long-term periods of detention ordered by courts, irregular detention and detention without charge, and conviction either overseas on charges connected to political activity or, *in absentia*, at home.

Stage 3

'End game', which includes a formal extradition to face torture and imprisonment, informal rendition or deportation often following release from detention, disappearance, serious attacks with an attempt to murder or disable, assassination attempt, assassination and other suspicious deaths.

The scale of transnational repression is measured on a 3-point ordinal scale according to these three stages. According to the stage of transnational repression experienced we have assigned a score between one and three, three being the most severe form of transnational repression experienced.

2.3 What is physical violence?

Physical violence captures forms of violence perpetrated, sanctioned, or ordered by agents of the state against an individual in political exile. The standard of evidence required to ascertain involvement by the state in violence is that of probability, as attributed in one or more of our authoritative sources. Physical violence includes:

1. Unknown – no information is provided on the type of physical violence experienced
2. Physical Attack – violence recorded by witnesses against a political exile short of a clear assassination attempt; exile remains alive after attack; an individual attempts to harm political exile through physical means.
3. Assassination Attempt – a clear attempt to take the life of a political exile
4. Assassination or Other Suspicious Death – the death of a political exile in suspicious or unknown circumstances including successful assassination attempts

The scale of physical violence is measured on a 4-point ordinal scale. According to the degree of violence experienced we have assigned a score between zero and four, four being the most severe form of physical violence experienced.

2.4 What is state of concern?

State of concern refers to the country of origin from where transnational repression is initiated. This country CANNOT be the country of ordinary residence of the person (as this person would not then be a political exile). This country MAY or MAY NOT be the country of citizenship of the exile. States of concern are one of the five post-Soviet Central Asian republics (Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan) according to the scope of the database.

2.5 What is gender?

In our database we distinguish between male and female. The cases do not include non-binary or unknown category since no persons in these categories appear in our final data.

There are no exiled transgender or related activist movements known to us from the public data. The reasons why non-binary persons do not appear as political exiles may reflect the conservative and authoritarian political contexts of Central Asian citizens both inside the region and in diaspora.

What is international arrest notice?

International arrest warrant refers to different types of arrest warrants issued by the state of concern to arrest or detain a wanted political exile. In CAPE dataset we distinguish between:

- Interpol Arrest Notice – evidence of an Interpol ‘red notice’ posted publicly (e.g. on the Interpol website), disseminated electronically, otherwise diffused, or provided to the exile, their counsel or a third party.
- Other Arrest Notice – evidence of another national or international arrest notice or warrant posted publicly, disseminated electronically, or provided to the exile, their counsel or a third party.
- Unknown – no information provided.

2.7 What is year?

Year in CAPE dataset corresponds to the highest stage incident level experienced by a political exile in a given year between 1990 and 2018 inclusive.

2.8 What is alleged affiliation?

Alleged affiliation is a sub-category of type of political exile. Alleged affiliation corresponds to an exile's publicly reported political affiliation activity or a presumed association claim made by his home state government. For reasons of method, we record publicly reported affiliation rather than self-declaration. We use publicly reported affiliation to place individuals into types of political exile.

2.9 What are countries of transit, asylum and/or residence?

Countries of transit refers to the countries through which an individual in political exiles passes in order to enter a country of destination. Multiple such countries may be recorded for an individual exile.

3. Coding and Methods

3.1 Variables and Codes

3.1.1 State of concern (Nominal)

The list of country codes from the Correlates of War project is taken directly from <http://www.correlatesofwar.org/data-sets/cow-country-codes>.

Format Name	Value Name	Value Label	Value (Numeric Code)
Country ID	Tajikistan	TAJ	702
	Uzbekistan	UZB	704
	Turkmenistan	TKM	701
	Kyrgyzstan	KYR	703
	Kazakhstan	KZK	705

3.1.2 Year

Format Name	Value Name	Value Label	Value (Numeric Code)
Year Incident Level	Year	Year	Year

3.1.3 Types of political exile (Nominal)

Format Name	Value Name	Value Label	Value (Numeric Code)
Category of Exile	Former regime insiders	Form_Reg_Insid	1
	Members of opposition political parties	Opposition	2
	Banned clerics and alleged religious extremists	Cler_Rel	3
	Independent journalists, academics and civil society activists	Jrn_CS	4
	Others	Others	5

Note:

In some instances, the exile could bear more than one characteristic. For example, the case of Mukhtar Ablyazov who was a former political insider and chairman of Bank Turan Alem and the co-

founder and a leader of the unregistered political party Democratic Choice of Kazakhstan. In such cases variables were recoded as follow:

Format Name	Value Name	Value Label	Value	Re-coded value
Category of Exiles	Former Regime Insiders/Opposition	FRI_Oppos	1/2	6
	BannedClerics/Civil society activist	Cler_CS	3/4	7
	Opposition/Clerics	Oppos_Cler	2/3	8
	Opposition/Civil Society	Oppos_CS	2/4	9

Format Name	Value Name	Value Label	Value (Numeric Code)
Stages of Transnational Repression	Stage 1	Notice	1
	Stage 2	Det_Arst	2
	Stage 3	Rend_Kidn	3

3.1.4 Stages of transnational repression (Ordinal)

3.1.5 Gender (Nominal)

Format Name	Value Name	Value Label	Value (Numeric Code)
Sex	Female	F_Gend	1
	Male	M_Gend	2

Format Name	Value Name	Value Label	Value (Numeric Code)
International Arrest Warrants	Interpol Arrest Notice	Interpol	1
	Other Arrest Notice	Other	2
	Unknown	Unknown	0

3.1.6 International arrest warrants (Nominal)

3.1.7 Physical Violence (Ordinal)

Format Name	Value Name	Value Label	Value (Numeric Code)
Physical Violence	Unknown	Unknown	0
	Physical Attack	Phys_Attck	1
	Assassination Attempt	Ass_Attempt	2
	Assassination or Suspicious Death	Ass_SDeath	3

3.1.8 Further notes

The dataset includes an individual ID number for each given political exile.

Individual affiliation and countries of transit have not been coded in the dataset.

Missing values: Values for missing data were coded as -99

3.2 Sources and Methods

Ten named 'authoritative' alongside secondary or 'non-authoritative' sources are used to produce raw data.

Using these sources is a two-stage process.

First, we search the following ten authoritative sources for relevant cases by state of concern

- European Court of Human Rights collection of 'Grand Chamber' judgements
- UN Human Rights country reports and special rapporteur reports
- Human Rights Watch country reports and special reports
- Amnesty International Urgent Actions and special reports
- US State Department Bureau of Democracy, Human Rights and Labor country reports
- Eurasianet (news) – search website
- Fair Trials International – search website
- Association for Human Rights in Central Asia – search website
- Norwegian Helsinki Committee – search website
- Radio Free Europe / Radio Liberty – search website

Second, once a name has been identified from one of these sources we will check the name in both Russian and English (and in common alternate spellings), going back over the 10 authoritative sources and any additional sources which may be used to complement the authoritative sources. The Interpol 'Red Notice' list is non-authoritative in the sense that research has demonstrated the presence of multiple lists where targeted political exiles may feature on non-public lists. In addition, red notices may be challenged by exiles, potentially removed, and on removal are not archived. These additional sources are 'secondary' in that they take place after a name has been identified in one of our 10 authoritative sources.

- Interpol 'Red Notice' list online
- Google search for other online and general press sources
- Any additional public sources known to the team or our reviewers

These sources are non-authoritative and will never be the sole basis of a case.

We may discover a new case from social media or referral by an expert, an exile themselves or their representative. However, we will never create an entry without going through stages 1 and 2.

3.3 Ethics

Ethical issues arise in using social media for data collection, including the collation of its publicly available information and researching Internet-based phenomena (Lunnay et al., 2015; Markham, A., & Buchanan, 2012)⁵. Our decision-making was guided by ethical principles of vulnerability, respect for persons, and beneficence.

⁵ Markham, A., & Buchanan, E. (2012). Ethical decision-making and internet research: Recommendations from the AoIR Ethics Working Committee (Version 2.0). Chicago: Association of Internet Researchers, Research Ethics Working Committee

According to Elgessem (2015, p. 26-28⁶), there are weaker grounds for requiring consent for non-private information that individuals themselves made available in public postings such as debate forums, online postings or in social media. In such instances, research may be conducted without obtaining consent even where subjects do not necessarily expect that their public information will be subject to a study. However, in instances where there is an expectation that people on the outside will not gain access (for example password requirements), then there would be a breach of the expectation to use the information in research without consent. This condition does not apply with regard to any of our data sources.

Given our choice, on ethical grounds, to include only public data in the database we have not requested permission from the persons named in the data. There may often be grounds, both ethical and legal (with regard to the General Data Protection Rule⁷) to inform the subject where their data has been processed, where the information about their person is sensitive, or where the person concerned is vulnerable. Moreover, we assess that any attempt to contact individuals may increase the risk that they may be identified by the state as concern as being in collaboration with a research project whose work has been used to criticise the very same state. Therefore, with regard to vulnerability, respect for persons, and beneficence, our judgment is not to specifically inform or seek consent for our use of their already-public data. However, if any individual would like their entry removed from or corrected on our database, they may contact us to request such removal or alterations. In addition, we would cooperate fully with regard to Freedom of Information requests with regard to correspondence on cases, previous versions of an individual entry and any revisions.

4. Production

4.1 Production and Analysis of the Dataset

Production of the dataset with its descriptive fields and analytical categories takes place across five stages.

4.1.1 Add new entries based on searching sources – see 3.2.

To add a new entry name in the database we start by examining the individual's case according to our definitions. The inclusion of a new name in the database is identified by a trained student working on CAPE. We identify a political exile from a variety of sources (see 3.2.), the information was collected from multilingual sources (Russian and English) for the purpose of data accuracy. If an individual was identified as a political exile in one source, we would seek to corroborate the information with another source. If cases where the sources contained discrepant information, research assistants were told to use the information from the most complete data source. We would only proceed with one source, if that source is a listed authoritative source (see 3.2.). For each new individual entry, we would write brief background description of the political exile case (suggested word limit of 100 words), check the individual categories and associated incidents based on publicly reported facts.

⁶ Elgesem, D. (2015). Consent and information: Ethical considerations when conducting research on social media. In H. Fossheim, & H. Ingierd (Eds.), *Internet research ethics* (pp. 14-34). Oslo: Cappelen Damm Akademisk.

⁷ Foitzik, P. (2019). Publicly available data under the GDPR: Main considerations. *International Association of Privacy Professionals (IAPP)*. Available at: <https://iapp.org/news/a/publicly-available-data-under-gdpr-main-considerations/>

4.1.2 Identify categories for each new entry's fields – see 3.1.

We would check the attributes of a political exile according to the assigned categories (see 3.1.3) and definitions (see section 2). If in doubt over a category, the entry was brought to the attention of supervisory project staff.

4.1.3 Identify incident stages for each new entry's fields – see 3.1.

We would check the incident stage level of a political exiles according to definition (see section (section 2.2 and 3.1.4). The CAPE database also includes the year of the most severe incident stage experienced by an individual. However, for some incidents, the date is missing. The field was marked blank for the missing incident date. A brief description was written for incident stage level experienced by an individual in political exile (suggested word limit 100). It's important to note that not all the cases go through the 3 stages. We only check and add the description for the stages that are given according to the collected information.

5.Verification

The verification process involved first correcting any data entry errors of which the student who originally entered the data was aware. The reviewer would review 10 of the original set of 20 cases assigned to the student. When systematic mistakes were found, verifier would review all of the student's sets of cases. Thus, in cases where systematic mistakes were found, all of the cases entered by that particular student were verified. After the verification stage, students will be notified of the errors made in the database. Students who made a significant number of mistakes were removed from the database work.

6.Update

Our ongoing efforts to validate the CAPE data have consisted of efforts to compare and review CAPE data to other open sources in many cases this involves, to go back to original sources to check for the accuracy of interpretations in the original data set. Improving the validity of the CAPE data is an ongoing project. The updating of the database proceeds in two stages:

In the first phase we update current entries based on searching all sources by name (in Russian and English, with common alternate spellings). The information is updated with a brief background description of person and associated incidents based on publicly reported facts in more than one source except when the source is a listed authoritative source.

In the second phase the student updates the categories and incidents stages for the entry fields according to the collected information. If in doubt over category, mark entry to be reviewed by the project supervisor.

6.1 Supervisor check

The updating process is reviewed by the project supervisor. The project supervisor edits descriptive text to ensure text is paraphrased and main sources are clearly referenced. Check consistency between text written and categories assigned. Raise concerns over data reliability or category assignment with project director as appropriate. Mark any outstanding doubts for review by external peer reviewer.

7. Review Process

The database is reviewed by peers prior to each new edition being released. These peers are practitioners or academic experts on particular countries or movements. The revisions and additions

to the codebook and interface are all made in the interest of increasing data entry reliability while decreasing data entry error. The review process takes place as follows:

7.1 Peer review to check, supplement and validate entries

Peer reviewers (country and regional experts) are contacted and sent a copy of the database or a sub-set of it, the codebook and review instructions. Each entry should be checked by at least one and preferably two external reviewers.

7.2 Revisions are made following review

Minor revisions are made to specific entries on advice of reviewers. Entries may be removed on occasions where the reliability of the data is questioned or where authoritative non-public sources suggest the case as reported in public may be significantly inaccurate⁸. New entries may be added according to the process beginning with step 4.1.1.

7.3 Consider new entries recommended by reviewers

New entries may be added according to the process beginning with step 4.1.1.

7.4 Internal review

The database is shared among the project team and advisory board with a final opportunity quality, reliability and omissions to be identified according to 4.2.1. Anomalies and trends are discussed.

7.5 Analysis and commentary

The analysis on the website is updated (State-of-concern summaries and profile cases) and special blogposts written and published as appropriate.

7.6 A new edition is released for publication

The CAPE database was released in three editions in 2016 (1.0), 2018 (2.0) and 2020 (3.0).

⁸ 'Significant inaccuracy' is defined as reported facts being incorrect. Non-significant inaccuracy would include minor omissions (such as further unreported places of exiles) which do not contradict what is publicly available.