# Key Officers of U.S. Foreign Service Posts: A Dataset of U.S. Diplomatic Representation, 1966-2017

David Lindsey* Matt Malis† Calvin Thrall‡

July 14, 2026

**Abstract**

Diplomats have long played a crucial role in facilitating interstate relations: they negotiate international treaties and agreements, facilitate commerce, settle disputes, and generally represent their home governments abroad. Yet, due to a lack of administrative data on diplomatic personnel, extant scholarship has studied almost exclusively the highest-ranking bilateral diplomats—ambassadors—while neglecting the vast majority of the foreign policy bureaucracy. We introduce KOFSP, a dataset of all individuals designated by the State Department as the "key" officers at American diplomatic posts from 1966-2017. We collect this data from digitized directories. These consistently capture the senior leadership at diplomatic posts and the mid-ranking officers leading major administrative subunits within posts. Over time, the State Department designated additional "key" positions and the coverage grows beyond this, encompassing about a quarter of all American diplomats in the later years. We use record-linking record-linking techniques to track over 35,000 unique diplomats throughout their careers. We also provide data on the name, location, and category of post at which each diplomat was serving, as well as information about the position that they occupied at the post. We validate this data through comparison to other sources. KOFSP stands to substantially advance the study of diplomacy, foreign policy, and bureaucratic career advancement.

*Associate Professor, Department of Political Science, Baruch College (david.lindsey@baruch.cuny.edu).
†Assistant Professor, Department of Political Science, Texas A&M University (malis@tamu.edu).
‡Assistant Professor, Department of Political Science, Columbia University (cthrall@columbia.edu).

## Background & Summary

National governments have long sought to simultaneously maintain large numbers of bilateral relationships with other governments, each of which requires constant management and attention; yet leaders themselves can only be in one place at a time. As a result, the vast majority of regularized interstate relations are carried out by diplomatic agents to whom governments have delegated the task of representing them abroad. While diplomats are most famously known as messengers—and, of course, delivering messages from their home government to their host government is one of diplomats' most important roles [1, 2]—their involvement in foreign affairs is far deeper and broader than this. Diplomats play critical roles in the formation of international treaties and agreements, both as the actors responsible for bringing both parties to the negotiating table and often as negotiators themselves [3]; they facilitate billions of dollars worth of trade and investment deals, acting as matchmakers and advocates for their national champions [4, 5, 6, 7, 8]; they are frequently the actors who step in to resolve international disputes, or to head off potential disputes [9, 10, 11]; they are also crucial collectors of information and intelligence about their host states, transmitting tens of thousands of communiques (or "cables") back home on a daily basis [12, 13, 14].

Due to their wide-ranging involvement in bilateral political and economic relationships, diplomats are an increasingly frequent topic of study for scholars of international relations, international law, and international economics. Yet, with exceedingly few exceptions [15, 16, 8], empirical studies of diplomats have focused exclusively on ambassadors [17, 18, 19, 9, 20, 4, 11, 21, 22, 23]. Ambassadors, as top-ranking diplomats and leaders of embassies, are clearly worthy of study. However, they constitute only a tiny fraction of all diplomatic personnel. Throughout the 2010s, for instance, the U.S. diplomatic corps consisted of around 13,000 to 14,000 officers; only around 170 of those (1.3%) were ambassadors (see Table 9 below).

One likely reason for the diplomacy literature's focus on ambassadors is data limi-

Table 1: KOFSP in comparison with other recent datasets of diplomatic representation.

| | Scope | | | | | |
|---|---|---|---|---|---|---|
| Dataset | Geographic | Temporal | Diplomatic | Names? | Unit | N |
| COW Diplomatic Exchange [25] | Global-Global | 1817-2006 | Relations-only | No | Dyad-half decade* | 310k |
| Diplometrics [26] | Global-Global | 1960-2022 | Relations-only | No | Dyad-year | 434k |
| GenDip [22] | Global-Global | 1968-2024 | COM-only | No | Dyad-decade* | 89k |
| Arias (2026) [24] | Global-UN | 1946-2019 | COM, Deputy | Yes | Dyad-year | 22k |
| Jost and Li (2025) [27] | China-Global | 1949-2023 | Broad** | Yes | Officer-position | 11k |
| Arias and Smith (2018) [19] | U.S.-Global | 1779-2014 | COM-only | Yes | Dyad-year | 15k |
| Thrall (2025) [8] | U.S.-Global | 1940-2015 | Broad*** | Yes | Officer-position | 8k |
| KOFSP | U.S.-Global | 1966-2017 | Broadest | Yes | Officer-position-QY | 499k |

*Note*: COM = Chief of Mission. "Relations-only" means the dataset only denotes the level of bilateral diplomatic relations (e.g. representation at the level of chargé d'affaires, minister, or ambassador) but no information on individual personnel. *Modal unit, though for some time periods may be more or less frequent. **Covers career histories for diplomats who reached top positions in China's Ministry of Foreign Affairs. ***Covers the sample of diplomats who participated in the Association for Diplomatic Studies and Training (ADST)'s Oral History Program.

tations. As Table 1 shows, the majority of extant datasets on diplomatic representation only contain information about diplomatic relations themselves or *chiefs of mission*, top diplomats who are typically ambassadors. These individuals have larger public profiles than other diplomats, allowing for easier data collection; focusing on chiefs of mission is a sensible choice for datasets that aim to maximize temporal and geographic coverage [24, 25, 22, 26]. However, the trade-off is that these datasets only allow researchers to observe a small fraction of the true variation in states' diplomatic presence abroad. For example, they contain no information on the size of the diplomatic mission or the number or location of diplomatic posts; no information about the substantive policy expertise of the rank and file diplomats in a given state; and no information about diplomats' career histories. In fact, none of the three largest extant datasets even include diplomats' names.

To fill this gap, this paper introduces the KEY OFFICERS OF U.S. FOREIGN SERVICE

POSTS (KOFSP) datasets. These data, collected by digitizing, cleaning, and merging State Department directories, contain information on all "key" diplomatic personnel at all U.S. diplomatic posts, at the quarterly level, from 1966 to 2017. KOFSP builds on existing work that has generated richer datasets of diplomatic representation by limiting geographic focus to one state (or organization)'s diplomatic relations [24, 19, 27, 8]; unlike existing data in this tradition, which either focus solely on top-ranking diplomats [24, 19] or were constructed retroactively from the career histories of selected diplomats [27, 8], KOFSP provides systematic records on a wide range of U.S. diplomats drawn from contemporaneous government documents. At approximately 500,000 diplomat-position-quarter-year observations, covering between 12% and 27% of the entire U.S. foreign service (see Table 9), the KOFSP data dramatically expand scholars' ability to observe and study diplomatic activity.

The KOFSP data contain information on the diplomatic positions held at each post, the identities of the holders, and various other diplomat- and post-level metadata at a fine-grained level. At the post level, these data provide measures of the number and kinds of U.S. diplomatic posts around the world, as well as proxies for the total number of officers at each post, and the subject matter expertise represented at each post. The longitudinal nature of the data allows researchers to track changes in the size, composition, and status of diplomatic posts over time. Further, we use a combination of automated and manual record-linking techniques to create unique identifiers for each diplomat in our data. This allows users to follow the careers of individual diplomats as they rotate from posting to posting, moving across countries and up through the ranks of the foreign service. Some users may want to aggregate the data to the country-year or country-quarter-year level; as such, we provide additional datasets at these levels of aggregation with several variables that describe both the types of diplomatic *posts* and the diplomatic *agents* that the U.S. had in place in a given host state at a given time.

The directories underlying the KOFSP data cover the "key" personnel at U.S. diplo-

matic missions, but there is some variation in the designation of "key" officers over time. The directories always include the senior leadership at posts and the the mid-level foreign service officers leading their major administrative subunits (sections). At the typical embassy, these two categories cover the 7 senior-most foreign service officers at any given point in time. Over time, the directories expand the definition of "key" personnel to include two other major categories: attachés from agencies other than the State Department present at diplomatic posts (e.g., military attachés) and personnel with operational responsibilities of particular interest (e.g., officers serving as equal employment opportunity liaisons). The underlying data for the first two categories are comprehensive, consistent, and highly accurate. That is, KOFSP can be used to consistently track the composition of the senior and mid-level leadership at each American diplomatic establishment for over fifty years. Coverage of the latter two categories is less consistent, and we suggest some caution in using KOFSP to study them.

There are several potential scholarly uses of the KOFSP dataset. First, international relations scholars have long been interested in quantifying diplomatic representation and recognition, both as a determinant of material international outcomes such as commercial exchange [28, 29] and war [30], and as a measure of more abstract concepts such as importance, status, prestige, or ambition on the global stage [31, 32, 33, 34]. Our data opens up avenues for further progress on these topics with far more fine-grained measures of the intensity and quality of bilateral diplomatic representation than has previously been available. Relatedly, the diplomacy literature's turn toward studying individual diplomats has yielded important insights into the inner workings of the "engine room" of international relations [35]; yet, as discussed above, the literature has been limited in its almost exclusive focus on the narrow set of diplomats who serve as the titular heads of embassies [17, 18, 19, 9, 20, 4, 11, 21, 22, 23] The KOFSP data allow researchers to widen their scholarly aperture to gain a more comprehensive understanding of the individuals who carry out the day-to-day work of international relations. Finally, the data will be

valuable for researchers interested in the political economy of bureaucracy more broadly. Large bodies of research have investigated the determinants of individual bureaucrats' career advancement, and the consequences of these personnel selection processes for bureaucratic performance across policy domains [36, 37, 38, 39, 40]. The KOFSP data provide a new empirical setting for testing longstanding theoretical questions in this literature.

# Methods

## Foreign Service Posts

Our data are drawn from a series of publications documenting the "key" officers of U.S. foreign service posts. We describe these below, but begin here with a brief overview of the substantive context of foreign service posts—the establishments staffed by diplomats that represent the U.S. government abroad.

American foreign representation is organized into *missions*. Each mission represents the United States to either a foreign government or an international organization and is led by a chief of mission, typically an ambassador, who holds broad authority over essentially all U.S. government civilian activities in the country [41]. The president appoints chiefs of mission, who may be either career diplomats selected by the State Department (and recommended to the White House), or "political" appointees selected by the White House directly. The remaining staff at American diplomatic missions are almost universally career government employees, selected through the personnel systems of their home agencies.

Missions consist of one or more *posts*—separate establishments that carry out diplomatic work. In the bilateral context, the principal post is nearly always an embassy, typically located in the host country's capital city. Bilateral missions often include additional subordinate posts, most commonly consulates in other significant cities, while missions to international organizations typically have only a single post.

Missions carry out a wide range of activity, reflecting the breadth of U.S. government interests overseas. Mission staff fall into three broad categories: the State Department's foreign service personnel; personnel from other federal agencies, who carry out specialized work related to those agencies' functions; and locally engaged staff hired directly by the mission, who typically carry out administrative, clerical, and support roles. Locally engaged staff make up the majority of personnel numerically, but U.S.-based staff from the first two categories carry out nearly all substantive foreign policy work and are the focus of our attention here.

Missions vary considerably in size and scope; the largest (such as the U.S. missions to Mexico or China) have hundreds of American diplomats, over a thousand total staff, and attachés from a dozen or more agencies. The smallest (e.g., the mission to Palau or to the International Civil Aviation Organization in Montreal) have only a handful of American diplomats and as few as a dozen total staff. Within missions, embassies are typically the largest and most complex posts. Subordinate posts vary considerably in size, from single-employee consular agencies to massive consulates general (CGs) in especially important cities, comparable in size and complexity to many embassies.

Embassies and large consulates follow a relatively standardized organizational scheme. Each post is led by a principal officer with a deputy as second-in-command. At an embassy, the principal officer is an ambassador, and the deputy is the deputy chief of mission (DCM). At a consulate general, the principal officer is a consul general and a deputy typically holds the title of deputy principal officer. Below the deputy, State Department personnel are organized into several sections, each under the leadership of a section chief. Personnel from other agencies generally report to the lead representative of the relevant agency. At the modal post, the State Department sections correspond to the five "cones" of the foreign service: consular, economic, management, political, and public diplomacy, plus a security section. Smaller posts may combine or omit some of these sections, while larger posts sometimes add additional ones for specialized functions, such as labor or

counter-narcotics policy. At embassies, these top officials are collectively known as the "country team."

State Department personnel at diplomatic posts fall into two major categories: *generalists* (known as foreign service officers, or FSOs) and *specialists* (formerly known as foreign service staff). Generalists carry out substantive diplomatic work in the five cones described above, primarily receive training on the job, and dominate the leadership ranks. Specialists provide support functions to enable diplomatic operations, ranging from medical services to information technology or construction engineering. Specialists are hired into specific roles, typically based on preexisting professional qualifications. Several agencies other than the State Department—including USAID, Agriculture, and Commerce—also use the foreign service personnel system for their overseas staff. Like their State Department counterparts, these officers are career diplomats and are eligible for promotion into overseas leadership roles (including ambassadorships). All other agencies send personnel abroad from outside the foreign service system, who are not eligible as careerists for top embassy leadership roles.

## The *Key Officers* Publications

Our underlying data come from a series of State Department publications originally released under the title *Key Officers of Foreign Service Posts: Guide for Businessmen*, with some variations in that title over time (see Figure 1). We use *Key Officers* to refer to all versions of these publications. These publications identify foreign service posts, provide contact information for them, and list the "key" officers of each such post. In its early existence, *Key Officers* coexisted with two other publications that supplied information on *all* American diplomats: the *Foreign Service List* and the State Department's *Biographic Register*. Outside researchers discovered that those publications could be analyzed to identify CIA officers serving under diplomatic [42], leading the State Department to discontinue them in 1976 [43].

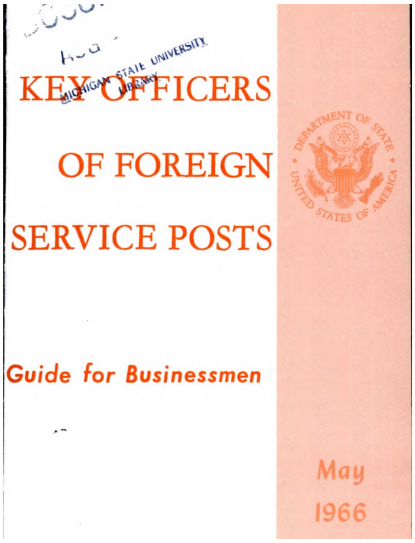
With the discontinuation of these publications, *Key Officers* became the only unrestricted source of information on American diplomatic staffing. This made it useful as an unclassified internal resource for the State Department and a a directory for other government agencies interacting with American diplomats [44]. Later editions eventually drop the reference to business in the title and provide information about personnel of no apparent interest to that community; for instance, liaison officers of the American Foreign Service Association (the professional organization representing foreign service officers) at each post.

While we have not been able to find any internal documentation of the underlying philosophy behind the compilation of *Key Officers*, an analysis of its composition strongly suggests that the "business" focus announced in the initial title is misleading. Rather, *Key Officers* is best understood as a general purpose directory of the people occupying leadership and other significant roles at diplomatic posts. The fact that the *State* Department publishes the directory reinforces this interpretation. The primary agency responsible for advancing American business interests overseas is the Foreign Commercial Service (FCS), housed within the Department of Commerce. The FCS publishes its own directory of commercial representatives abroad, which is both more detailed and more useful than *Key Officers* for business purposes, currently covering all overseas commercial diplomacy personnel and supplying direct contact information for them. In contrast, *Key Officers* consistently documents embassy staff with no commercial or business responsibilities and provides no specific contact information for the people it identifies.

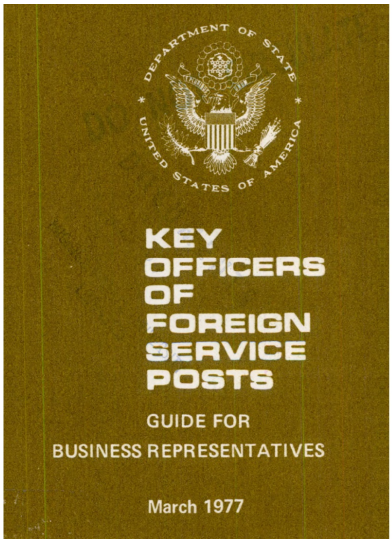
The specific positions identified by the State Department as "key" for the purposes of the publications have changed over time. Generally, we can group these into four categories: the senior leadership of each post (its principal officer and deputy); the middle-management layer of State Department officials running major administrative subunits (sections) of the post; the lead attachés from other agencies present at the post; and other personnel carrying out functions of particular interest. The earlier directories cover the

Figure 1: *Key Officers* publications over the years.

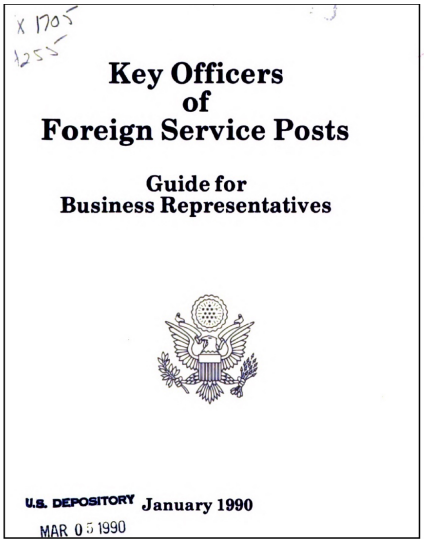
(a) May 1966



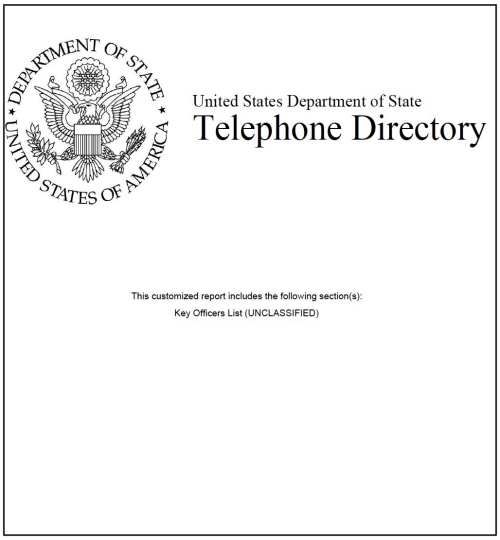

(b) March 1977




(c) January 1990




(d) April 2017




first two categories (i.e., effectively mapping out the foreign service leadership of American posts) with more personnel from the latter two groups added over time. With very rare exceptions, the directories identify a *single* person for each role or function of interest. At a large post, for example, there might be dozens of people employed to handle consular affairs, but only the officer leading the consular section (and managing those subordinates) will be listed in *Key Officers*.

The State Department primarily compiled *Key Officers* by periodically soliciting the relevant information directly from posts rather than updating it from centralized personnel files or other sources [45]. Consequently, there is no standardization in the ways that names are reported: nicknames are often (but inconsistently) used in place of proper names, components of names may be abbreviated or dropped, and obvious misspellings or typos are relatively common. Officers may also be listed before their arrival or after their departure, sometimes leading a single officer to be simultaneously listed in multiple locations. This makes it difficult (perhaps deliberately so for counterintelligence reasons) to construct persistent identifiers for officers over time. In contrast to the discontinued publications, for example, *Key Officers* contains no index or other resource organized by officer name. Prior to digitization, the only way to use *Key Officers* to locate an officer by name over time would have been to sequentially search through each page of each edition of the entire publication.

## Assembling the Key Officers Dataset

Our process for assembling the Key Officers (KOFSP) dataset began by obtaining copies of the publications. For the period from 1966 to 1998, the directories were published physically. We downloaded scanned and digitized copies of these publications from HathiTrust. From 1998 onward, the State Department published the directories online. We downloaded these through the Internet Archive.

In the earliest years, updates were published on a quarter-yearly basis. Publications became less frequent over time, reduced to once or twice per year by the 1990s. In the early years of the online-only publications, the Internet Archive's coverage is relatively sparse, with only seven versions of the directories available over the years 1999–2003. From 2004 onward, the directories were updated on a seemingly ad-hoc basis, but with at least quarterly frequency in most years. We collected the first version available in each quarter, for consistency with the print-era publications.

Figure 2: Sample Entry from "Key Officers" Publication

**MOROCCO**

**Rabat (E),** 2 Ave. de Marrakech • PSC 74, Box 003, APO AE 09718, Tel [212] (7) 76-22-65, Fax 76-56-61, afterhours Tel 76-96-39, Telex 31005M; USAID Fax 70-79-30; USIS Fax 75-08-63

| | |
|---|---|
| AMB: | Marc C. Ginsberg |
| AMB SEC: | Kelli Adams |
| DCM: | Gary S. Usrey |
| POL: | Joseph Mussomeli |
| ECO: | Alexandra M. Sundquist |
| POL/LAB: | Elizabeth Lee Martinez (resident in Casablanca) |
| CON: | Andrea Levin |
| ADM: | Richard E. Kramer |
| RSO: | Lawrence H. Liptak |
| PAO: | James L. Bullock |
| COM: | (Vacant) |
| IMO: | Nicodemo F. Romeo |
| ISO: | Lucille Smithson |
| AGR: | Quintin Gray |
| AID: | Michael Farbman |
| DAO: | Col. Alan J. Tinder, USAF |
| ODC: | Col. Stephen Fenton, USAF |
| IRS: | Frederick D. Pablo (resident in Paris) |
| FAA: | Tony Fazio (resident in Paris) |

**Casablanca (CG),** 8 Blvd. Moulay Youssef • PSC 74, Box 24, APO AE 09718 (CAS), Tel [212] (2) 264-550, Fax 204-127; COM Fax 220-259; USIS Tel. [212] (2) 221-460, Fax [212] (2) 299-136; Duty officer's cellular Tel. no 13-4065

| | |
|---|---|
| CG: | Marcia Bernicat |
| POL/LAB: | Elizabeth Lee Martinez |
| ECO: | Daniel K. Balzer |
| COM: | Albert Nahas |
| CON: | Shelley S. Midura |
| ADM: | Sylvia Nasri |
| IPO: | Raymond Harger |
| BPAO: | Philip A. Frayne |

With the directories downloaded, our first objective was simply to capture the raw data from the directories faithfully in a standardized format. For the print publications from HathiTrust, we used a commercial data-entry firm to convert the inconsistently formatted original scans into consistently formatted spreadsheets. For the online publications from the Internet Archive, we parsed the digital copies into similar spreadsheets.

Figure 2 shows an example entry in the *Key Officers* publication for Morocco in June 1997. We extract the following information directly from the publications:

- The name of the country in which the post is located (at the top, in all caps);
- The name of the post (bolded, typically the name of a city, here "Rabat" and "Casablanca");
- The type of post (a code following the post name in parentheses, here "E" for embassy and "CG" for consulate general);
- The position for each officer (given here as an abbreviation in all caps);

- The name of the officer occupying each position (which may be vacant, as in the COM position at Rabat);
- Any parenthetical note indicating that the officer resides in some other place (such as "resident in Paris" for the IRS and FAA entries in Rabat).

We collect this information for 156 unique quarterly directories issued between 1966 and 2017. We note here that the data end in 2017 simply because that was the year in which the authors began the project; *Key Officers* directories continue to be published, and we discuss the prospects for updating the data in the Data Availability section below. In total, the data contain approximately 500,000 officer-position-quarter-year observations. We specify that the unit of observation is the officer-*position*-quarter-year, rather than the officer-quarter-year, because some officers hold multiple positions in the same quarter-year. This can occur when the same officer is listed as simultaneously occupying two positions at a given post, or when the officer is simultaneously listed at two different posts; we discuss the latter case in more depth in our discussion of the `residence_location` variable.

## Cleaning the Data

After collecting and concatenating the data from the Key Officers directories, we performed a few basic cleaning tasks. First, we took a number of steps to prepare the officer names for record linking: we removed all suffixes (e.g. "Jr.", "III") and military ranks/titles (e.g. "Col.", "Maj") from officer names and stored them in new variables; removed all prefixes, such as "Ambassador;" removed nicknames and middle initials and stored them in new variables (for internal linking use only; these are not included in the published datasets); split apart observations that listed multiple officer names for the same position (e.g. "Paul Simon and Art Garfunkel"); and fixed general spacing, capitalization, and other typographical errors. This process created the `officer_name` variable,

while we also retained the uncleaned names in the `officer_name_raw` variable.

We also performed some light cleaning of position names, primarily by fixing clear typographical errors and harmonizing position naming conventions across time. We then used a combination of (1) position directories contained in the key officers documents themselves, (2) information from the Foreign Affairs Manual (https://fam.state.gov), and (3) other secondary sources [46] to determine the tasks associated with each position type and which category of officer would hold each given position. We code this information in the `position_category` variable, which we discuss in greater depth in the Creating Other Variables section below. As will be detailed in the following section, we use these position categories to help refine our record-linking procedure.

Finally, we also perform some light cleaning and harmonization of post, post type, city, and country names. Our most substantive cleaning decision in this area concerns the `country` variable; to foster comparability over time, we use the `countrycode` R package [47] to convert all country names to their standardized contemporary versions. For example, while the country currently known as the Democratic Republic of the Congo was named Zaire from 1971 to 1997 (and appears as such in the directories), we code it as the Democratic Republic of the Congo for the entirety of the sample. Users wishing to find the contemporaneous names under which each country appeared in the Key Officers directories may access the directory PDF files from our Dataverse page [48].

We have generally preserved the State Department's ontology of "countries" as represented in the directories. These are not the same as sovereign states recognized by the United States. Rather, the State Department frequently lists dependent territories as their own separate entities in *Key Officers*. For example, Hong Kong is continuously listed as a separate entity, rather than being listed as part of the entry for the relevant sovereigns (first the United Kingdom, and later China); thus, we list Hong Kong as its own country. In some cases, the directories' classifications of countries vary over time. In such cases, we adopt the most recent treatment of the post and extend this backward over time. For ex-

ample, the American consulate in Ponta Delgada in Portugal's Azores islands is generally listed in the Portugal entry. In 2004 and 2005, this same consulate is instead listed under a separate "Azores" entry (despite no underlying change in the status of the territory); we identify Portugal as the country throughout these years.

## Creating Unique Officer and Spell Identifiers

Our process for linking officers over time involves both an initial algorithmic process and then extensive manual cleaning and validation. Our first step is the identification of "spells"—that is, a set of officer-position-quarter-year observations that refer to a single officer's uninterrupted tour at a given post. We then link together spells to uniquely identify officers.

The basic process for identifying spells proceeds as follows:

1. Presumptively link together cases where an officer is present in the same country with an exact first name-last name match (within 2 years of previous appearance) or a fuzzy first name-last name match with a Levenshtein edit distance of less than 7% of the string length (within six months of the previous appearance).

2. When the first step generates gaps (i.e., the officer is not continuously present between their first and last linked appearance), relax the maximum fuzzy matching distance to 30% but additionally require an exact match on the officer's position. Manually verify candidate matches generated by this method.

We then link together spells as identifying the same officer based on the following criteria:

1. An exact first name-last name match within 10 years of the previous appearance.

2. When the first name-last name match yields multiple possibilities, we further refine to those with a matching middle initial.

3. We then apply fuzzy matching with a maximum relative first-name last name distance of 10% of the string length, further refined as described below.

Next, we identify factors that disqualify certain matches. When names are matched exactly, any disqualifier leads to a manual review. When names initially matched only approximately, then matches with a disqualifier are discarded. We use the following criteria:

1. Matches that imply that the same person was in two places at once for more than a single quarter.

2. Matches that imply that an officer transitioned between incompatible career categories based on their position category; specifically, we look for cases where an officer moved between a foreign service position and an external attaché position, given that this is generally not possible under the relevant personnel rules.

3. Failed middle name or initial matches (i.e., cases where the first and last names match, but the middle names do not).

From this basic starting point, we began an extended process of manual match refinement.

**Manual Refinement: False Positives**

After the initial round of automated record linking, we begin our process of manual match refinement by looking for false positives: instances in which we have incorrectly combined two or more different officers into a single officer ID. To do so, we first identify every single unique officer ID that is associated with two or more unique names (e.g. "Joe Brown" and "Joseph Brown"). These are the most likely source of false positives in our data, as they represent cases in which our linking algorithm decided that officer-position-quarter-years associated with different names were in fact referring to the same individual. We conduct manual refinement to ensure that these decisions were justified.

We then manually inspect all 893 officer IDs associated with three or more unique names. Manual inspection consists primarily of visualizing each officer's career trajectory and determining whether or not the officer ID has accidentally combined multiple different officers. When possible, this process was aided by sources such as CVs, obituaries, personal webpages, and other forms of "ground truth" that allow us to distinguish different officers' careers. When such documents could not be found, we relied on our own understanding of typical diplomatic career paths to determine whether and how one officer should be separated into multiple officers.

Next, out of the 3,325 officer IDs that contained two unique names, we again automated the selection of likely false positives for manual inspection. We applied the following criteria to identify potential false positives for inspection:

- If the officer has only held one position category throughout their career, they are **not** a likely false positive; it is unlikely that two or more different officers shared common names and also held common positions.

- If the officer has only ever served at a single post, they are **not** a likely false positive; it is unlikely that two or more different officers shared common names and also served at the same post.

- If the officer has held multiple different attaché positions throughout their career, they **are** a likely false positive; it would be quite unusual for a single officer to represent multiple different executive agencies at U.S. diplomatic posts over the course of their career.

- If the officer has held two or more of the positions [attaché, specialist, chief of mission] throughout their career, they **are** a likely false positive; these are not positions that often appear in the same career trajectory as one another, so if we observe one officer who has held more than one of them we may suspect that we have incorrectly combined multiple officers into one officer ID.

- If the officer's unique names are sufficiently different (in terms of string distance), they **are** a likely false positive; while "John Davis" and "John Davies" might plausibly refer to the same officer, it is quite unlikely that "John Davis" and "Tom Davis" do.

Of the 3,325 officer IDs with two unique names, this procedure identified 819 IDs meriting further manual inspection. After repeating our manual inspection process with these officers, we concluded our systematic search for false positive linkages.

**Manual Refinement: False Negatives**

Next, we began searching for false negatives: cases in which we incorrectly separated one officer's career into multiple unique officer IDs. There are many reasons why we might fail to link an officer to herself: inconsistently used nicknames, multiple spellings of the same name, name changes, unusual career trajectories, and so on. Thus, we take multiple steps to address false negatives in our data.

First, we manually investigate all cases in which we have assigned multiple unique officer IDs to the same exact name (first and last). In many cases, it is clear upon manual inspection that all IDs refer to the same individual, and that the split was due to the imperfection of our initial linking algorithm. For example, Michael Hornblow (officer ID 1910) was incorrectly split into two unique officers because there is an 11 year gap between two of his consecutive appearances in the Key Officers directories, and we initially set a 10-year maximum gap for linking records to one another. In other cases it is less clear, and we again must rely on any corroborating public documents and/or our own understanding of diplomatic career paths to assess whether or not all observations associated with a particular name refer to the same individual.

Next, false negatives can also occur when we fail to link an officer to themselves because their name appeared in multiple different spellings in various copies of the Key Officers directories; while we use fuzzy matching to allow for some difference between

spellings, large differences will fail to link. As a result, we perform manual inspection among a subsample of officers for whom we believe that false negatives are especially likely: those with common first names that also have common short versions or nicknames that are significantly different than their full versions (such as Michael/Mike). For eight first names that fit this category—John/Jon/Johnathan/Jonathan; (2) William/Will/Bill; (3) David/Dave; (4) Michael/Mike; (5) Thomas/Tom/Tommy; (6) Joseph/Joe/Joey; (7) Stephen/Steven/Steve; (8) Christopher/Chris—we perform manual inspection by (1) identifying all unique officers with any version of each first name; (2) sorting them alphabetically by last name; (3) determining whether officers with the same last name and different versions of the same first name (e.g. "Bill Anderson" and "William Anderson") should be combined. Again, we make these determinations based on corroborating public documents if possible, and otherwise based on our own judgment.

Our automated linking process is also likely to produce false negatives for officers who occasionally appear with their first and middle names spelled out (e.g. "David Lee Roth") but also occasionally appear with at least one of their names as an initial only (e.g. "D. Lee Roth", "D. L. Roth", etc). To address this issue, we first compiled a list of every officer name that takes one of the following forms: (1) [first name] [middle name / second name] [last name]; (2) [first initial] [middle name / second name] [last name]; (3) [first initial] [middle initial] [last name]. We then created a function to search the full dataset for other permutations of these names—for example, matching "David Lee Roth" to "David Roth" and "D. L. Roth"—and then performed our standard manual evaluation of whether or not all matches should be combined.

Finally, false negatives can be created if an officer *changes* their name entirely for any reason. Marriage and divorce are likely the most common sources of name changes in our data: most commonly, either one spouse will take the other's last name or both spouses will adopt a new hyphenated last name (and in the case of divorce, the opposite). Unfortunately, we simply lack sufficient corroborating evidence to systematically identify

Table 2: Position categories in the KOFSP data.

| Position Category | Description | Position Examples |
|---|---|---|
| Generalist | State Department officers working on general diplomacy | Consular Officer, Political Officer, Economic Officer |
| Specialist | State Department personnel hired for specific (mainly administrative) skills | Administrative Assistant, Facilities Management, Information Security |
| Nonstate FSO | Foreign service officers from other executive agencies | Foreign Commercial Service, Foreign Agricultural Service |
| External | Non-foreign service personnel from other agencies | DHS Attaché, FAA Attaché, Defense Attaché |
| Undetermined FSO | Positions that can be held by multiple types of FSO | Community Liason Officer, AFSA Representative |
| Unclear | We cannot determine the type of officer who would hold these positions | ICASS Chair, REP, Advisor, Counselor, other ambiguous acronyms |

cases in which an officer has adopted (or relinquished) their spouse's last name but remains the same individual. This is a limitation of the data. However, for all officers with hyphenated last names (e.g. "Julia Louis-Dreyfus"), we search the full dataset for any officers with the same first name and only one of the last names (e.g. "Julia Louis" and "Julia Dreyfus") and repeat our manual inspection procedure to identify whether these individuals should be combined.

## Creating Other Variables

We construct several other variables which are not directly drawn from information presented in the Key Officer publications. These variables include `position_category`, `position_agg`, `senior_officer`, `political_appointee`, `residence_location`, and `residence_certainty`. This section describes the construction of each of these variables.

**`position_category`:** Our best assessment of the personnel category that the officer falls within, based on cross-referencing the officer position abbreviations in the *Key Officers* publications with information from other State Department resources. This variable takes on the following values: Generalist (44% of observations), Specialist (23%), Nonstate FSO (12%), External (12%), Undetermined FSO (7%), and Unclear (2%). Generalists and Specialists are the two broad personnel categories within the Foreign Service. "Nonstate FSO" refers to Foreign Service Officers housed in departments or agencies outside of the State Department (see discussion in the Foreign Service Posts section above). "External" refers to non-FSO employees from other executive branch agencies (e.g., IRS, DEA, FAA, or military attaches). "Undetermined FSO" denotes that the position is typically occupied by a member of the foreign service, but we cannot determine which subcategory. "Unclear" is a residual label for officer positions whose position category we cannot narrow down at all. Table 2 lists the six position categories, with descriptions and examples of each.

We note that the coding of the `position_category` variable is exclusively based on the entry in question, representing only the information available from the officer's position. For example, if an officer is listed both in a position classified as "unclear" and separately in a "generalist" position, we do not fill in the "generalist" classification into the entry that is otherwise unclear. As a result, around 15% of officers (5,497 out of 35,477) occupy positions of different position categories over the course of their careers.

Table 3 reports the breakdown of position categories across post types. For clarity of presentation, we group "Nonstate FSO" together with "External" officers, and group "Unspecified FSO" and "Unclear" with "Specialists". The table shows that the vast majority (75%) of our officer-quarter-year observations are posted at bilateral embassies, rather than other post types. Though embassies only constitute a small portion of the posts in our data (36%), embassies are almost always the largest posts in their respective countries.

Table 3: Breakdown of Position Categories by Post Type

| | Generalist | Specialist/ Others | External | Unique Posts | Total Officer-Position-Quarter-Years |
|---|---|---|---|---|---|
| Overall | 0.46 | 0.32 | 0.22 | 478 | 457,746 |
| Embassies | 0.43 | 0.34 | 0.23 | 193 | 342,621 |
| Other Posts | 0.55 | 0.27 | 0.18 | 340 | 115,125 |

*Note*: Values in the bottom two rows of the "Unique Posts" column do not sum to 478 due to posts changing status over time (e.g. from consulate general to embassy). Number of observations in the table (457,746) is less than the total number of observations in the dataset (498,503) because the table drops observations of `residence_location` = "Unclear" (see below).

**`position_agg`:** A grouping system for officer positions (short for "aggregate position") which is more disaggregated than the `position_category` variable, taking on 29 distinct values. Whereas `position_category` is assigned based on an officer's personnel system or home agency, `position_agg` is assigned based on the functional role of the position.

Note that `position_agg` values are mostly, but *not* completely, nested within `position_category` values: that is, 20 of the 29 `position_agg` values fall within a unique `position_category` value. For instance, all "Treasury/IRS" officers are external, and all "Economic" officers are generalists; but `position_agg` = "Law Enforcement" combines both generalists leading narcotics and law enforcement (NAS/INL) sections and external attachés from law enforcement agencies such as the FBI and DEA. Additionally, our coding accounts for organizational changes throughout the time period. For instance, the Foreign Commercial Service moved from the State Department to the Department of Commerce in 1980, so their `position_category` variable changes from "Generalist" to "Nonstate FSO". Likewise, Public Diplomacy officers are coded as "Nonstate FSO" prior to the incorporation of the U.S. Information Agency (USIA) into the State Department in 1999, and "Generalist" thereafter.

**`senior_officer`:** An indicator for the highest-ranking officer at each embassy or CG. Specifically, within each embassy, we code this variable 1 for the officer holding the highest of the following positions (in descending order): Ambassador, Chargé d'Affaires

(CHG), Chief of Mission (CM), Principal Officer (PO), or Deputy Chief of Mission (DCM). Within a CG, the ordering of positions is: Consul General (CG), Principal Officer (PO), Deputy Principal Officer (DPO). At nearly all embassies and CGs, we identify a single senior officer, and code 0 for all other officers at these posts; the variable is coded NA for all officers at all other post types.

**`political_appointee`:** A binary variable denoting whether an officer is identified as a "Non-career appointee" by the State Department Office of the Historian's Principal Officers and Chiefs of Mission Database [49]. The Historian database records all U.S. ambassadors, and other State Department appointees with the rank of ambassador and above (such as U.S. representatives to international organizations, and Washington-based positions at and above the level of Assistant Secretary of State). Each officer in the Historian database is identified as either a "Career Foreign Service Officer" or a "Non-career appointee". We linked officers from Principal Officers to their corresponding entries in Key Officers, and coded `political_appointee` = 0 for those listed in Principal Officers as "Career Foreign Service Officer", 1 for those listed as "Non-career appointee", and NA for everyone else (leaving the overwhelming majority of observations—454,253, or 91%—coded NA, as most key officers never reach a position at or above the level of ambassador). Political appointees are almost exclusively ambassadors, high-level officers at IO missions, or principal officers at other high-profile posts (such as the U.S. "Liaison Office" in Beijing in the mid-1970s); in only a handful of cases do political appointees appear in other embassy positions such as DCM or as USAID mission director.

**`residence_location` and `residence_certainty`:** These variables denote where an officer is physically posted, if different from the post at which they are listed. Our coding of `residence_location` begins with the information reported explicitly in the Key Officer publications. In Figure 2, for instance, we see Elizabeth Lee Martinez listed twice in the June 1997 publication: first at the embassy in Rabat, with the parenthetical "(resident in

Casablanca)", and separately at the CG in Casablanca. In this case, the `residence_location` variable is coded as "resident in Casablanca" for the Rabat observation, and left NA for the Casablanca observation.

The `residence_certainty` variable takes one of four values: "post_known", "country_known", "in_transition", and "unclear". Both Elizabeth Lee Martinez observations in June 1997 are coded as "post_known". (We explain the other values below.)

We further modify these variables to account for cases in which the same officer appears in multiple places in the same quarter-year, without an explicit acknowledgment in (at least) one of the listings that they are resident elsewhere. For example, Jeff Wagner (officer ID: 17972) is listed as DEA attaché in both Port Louis, Mauritius and Pretoria, South Africa from 2004q4 through 2006q3. Because we have no way to determine at which post he actually resides, we code all observations across the two posts over this period as `residence_location` = "unclear" and `residence_certainty` = "unclear". For use convenience, we also include a binary variable `officer_is_at_post`, which takes a value of 0 when both `residence_location` and `residence_certainty` are "unclear" (the two always coincide), and 1 otherwise.

There are two special cases of multiply-listed officers which we handle differently. First, when there is an overlap for only one quarter (at the end of one "spell" and the start of another), we consider that to be a transitional quarter. We assume that the officer is in fact resident at both locations during some portion of the transition quarter. For example, Jeri Guthrie-Corn (officer ID: 8719) is listed as DCM in Bucharest from 2008q4 through 2011q4, and as DCM at the U.S. Mission to the OECD in Paris beginning in 2011q4. In this case, we code `residence_certainty` = "in_transition" for both observations in 2011q4 (and "post_known" for all other surrounding observations), and leave `residence_location` as NA for the entirety of both spells.

Second, when an officer is simultaneously listed at an embassy and another subordinate post within the same country (but with no indication in either listing that she is

Table 4: KOFSP datasets and their contents.

| Filename | Aggregation Level | Rows | Variables |
|---|---|---|---|
| KOFSP_individual_v1.1.csv | Officer-position-quarter-year | 498,503 | 26 |
| KOFSP_post_quarter_v1.1 | Post-quarter-year | 42,719 | 24 |
| KOFSP_country_quarter_v1.1.csv | Country-quarter-year | 24,100 | 26 |
| KOFSP_country_year_v1.1.csv | Country-year | 7,961 | 23 |

resident elsewhere), we assume that the officer is actually posted at the embassy. For example, Thomas McGrath (officer ID: 5394) is listed as Regional Security Officer (RSO) simultaneously at the embassy in Bonn (Federal Republic of Germany) and the CG in Dusseldorf in 1984q4 and 1985q1. For all four observations, we code `residence_certainty` = "country_known"; for the Dusseldorf observations, we code `residence_location` = "BONN", and we leave `residence_location` as NA for the Bonn observations.

# Data Record

All KOFSP datasets, as well as the source directories from which we constructed them, are freely available for download on our Harvard Dataverse site (https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/3IPF8Y) [48]. Table 4 describes the four datasets available for download: one at the officer-position-quarter-year level, which is the level at which the data was reported in the directories and which contains officer-level metadata; one at the post-quarter-year level, which aggregates key variables at the level of the post-quarter-year; one at the country-quarter-year level, which aggregates key variables across all posts in a given country in a given quarter-year; and one at the country-year level, which aggregates key variables across all posts and quarters in a given country-year. Most variables in the aggregated datasets are intuitively named, and a full description of each of them can be found in the codebook on our Dataverse site [48].

# Technical Validation

We perform three sets of validation exercises. First, we validate our record linkage via comparison of our data to a separate sample of verified diplomatic career histories. Second, to demonstrate that the number of key officers at a given post is a strong proxy for the number of total U.S. personnel at that post, we compare our data to a sample of governmental audit reports that contain information on the latter. Finally, we use the same audit reports to validate the underlying *accuracy* of the *Key Officers* directories—that is, to demonstrate that the officers listed in the directories are indeed the officers actually serving at their respective posts.

## Technical Validation 1: Validating Career Histories via Oral Histories

To validate our process of linking entries together to create unique identifiers for each officer, we rely on an additional data resource: oral history interviews published by the Association for Diplomatic Studies and Training (ADST). The ADST oral history project has interviewed over 2,500 individuals associated with American diplomacy. The format and objective of these interviews vary, but many of them supply comprehensive career histories for diplomats, providing the ground truth for a particular officer's assignments. We randomly sampled 200 of these interviews and manually reviewed them, yielding a sample for analysis of 110 officers with a complete career history who held at least one diplomatic posting during the years covered by our dataset. The interviews supply information on a total of 530 diplomatic postings held by these officers. 291 of the 530 postings correspond to a listing in *Key Officers*; the remainder cover positions that do not appear in Key Officers (typically because the position in question was too junior). A total of 101 officers from the ADST sample appeared at least once in *Key Officers*, and we focus on these below.

For each officer, we manually review each position to determine if the officer has been

linked correctly over the course of their career. We are able to systematically detect both over-linking (cases where we have linked an officer to an entry that actually refers to someone else) and under-linking (cases where we failed to link an officer to a posting they did hold) as well as to evaluate any other relevant patterns. In conducting this process, we also discovered a third class of outcomes—underlying data errors where the officer is linked correctly to an entry, but the underlying entry in the *Key Officers* publications was in error.

Table 5 shows the error rate in the data at both the position-level (i.e., is the particular position correctly linked?) and the officer-level (i.e., aggregated up to the officer level, are there any errors in any of the positions?).

Table 5: Comparing KOFSP to ADST: error rates.

| Category | Number of Officers | Number of Positions |
|---|---|---|
| No Errors | 88 | 284 |
| Under-linking Error | 7 | 7 |
| Over-linking Error | 2 | 3 |
| Underlying Data Error | 4 | 4 |

At the aggregate level, we find no errors for 88 of the 101 officers (87%). At the position-level, we find no errors in 284 of 298 positions (95%; note that the denominator is 298 here rather than 291 because over-linking increases the number of positions we are analyzing relative to the number of positions our officers actually occupied). All of the detected under-links involve a name variant in the underlying data (nicknames, typos, or exclusive use of a first or middle name across separate postings). For instance, we failed to link appearances of William S. Rhodes (officer ID: 12151) with appearances of Stacey [sic] Rhodes (officer ID: 14614); this officer's full name is William Stacy Rhodes, but he is never listed in Key Officers under his full name.

Both over-links involve officers who coincidentally match another officer with exactly the same first and last names and an apparent career trajectory that is sensible. So far

as we can tell, all four of the underlying data errors involve cases where an officer had been selected for an assignment, was listed in the *Key Officers* publication in anticipation of arrival, but then was reassigned at the last minute. Each of these four listings lasts for only a single quarter. In two of the four cases, we can explicitly confirm this sequence of events based on the oral history interview. In the other two, we believe the same sequence is highly likely based on the surrounding context.

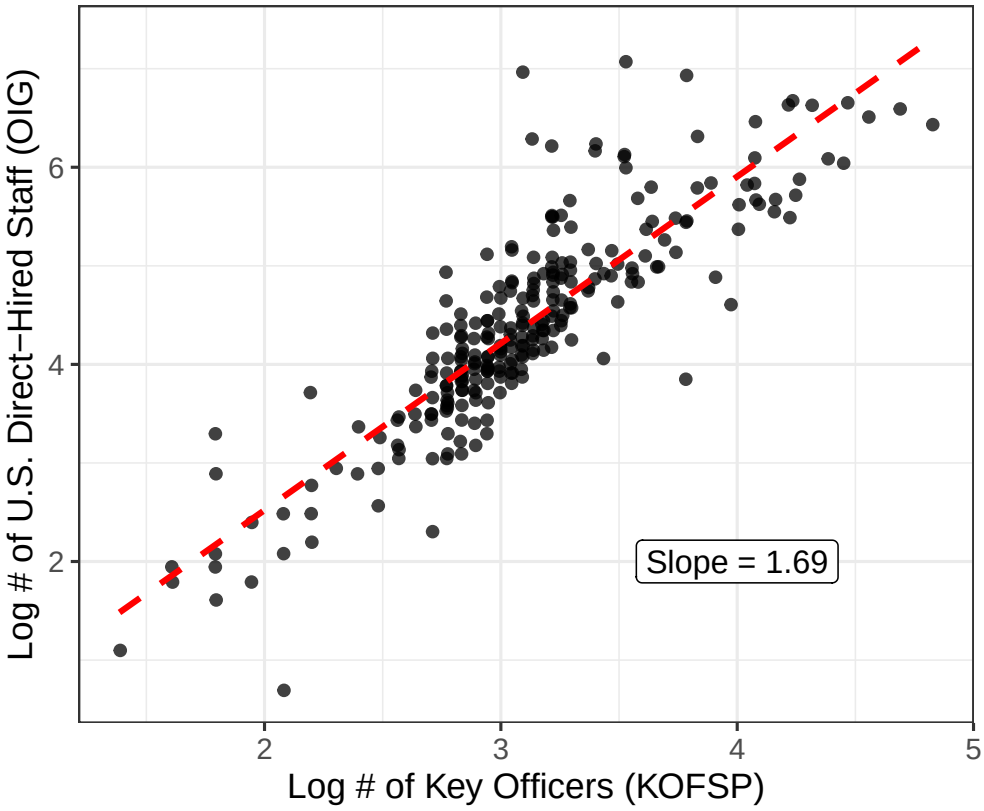
Extrapolation from the error rate in this sample to the broader population requires some caution. Officers who have chosen to complete an ADST oral history interview undoubtedly differ in various ways from those who have not. Most obviously, the ADST sample consists almost exclusively of retirees, so we are biased away from the most recent years of the KOFSP dataset. The ADST interviews also over-sample officers with longer careers who achieved higher levels of seniority. Nonetheless, we see no particular reason to believe that these attributes make it either significantly easier or harder to carry out the linking process.

## Technical Validation 2: Key Officers as a Proxy for Embassy Size

To validate the use of the key officers data as a proxy for embassy size, we rely on inspection reports released by the State Department's Office of Inspector General (OIG). OIG periodically evaluates operations at American diplomatic missions, and its inspection reports routinely include information on staffing levels. This information is not otherwise publicly available. We have gathered all inspection reports for American bilateral diplomatic missions (i.e., excluding missions to international organizations) available on OIG's website that report staffing information for a total of 265 reports released between 2002 and 2017.

We are primarily interested in the ability of our measure to proxy for the number of diplomats in a country (i.e., staff hired by U.S. government agencies in the United States and then sent abroad). The OIG reports provide both this information and data on the

Figure 3: Key Officers and Total U.S.-based Officers




total staffing levels at missions inclusive of locally engaged staff—that is, host country nationals hired by a mission to assist with administrative work (e.g., security guards, drivers, clerical staff, etc.).

In Figure 3 we plot the natural logarithm of the number of distinct key officers at a mission (from KOFSP) against the natural logarithm of the total number of American direct-hire staff at that mission from the OIG reports. The correlation between the two logged variables is 0.86. The relationship between the logged number of distinct key officers and the logged total staff level (inclusive of locally engaged staff) is slightly weaker but still quite strong at 0.78. In sum, the number of key officers is a highly effective proxy for embassy size.

We note that it is possible, though uncommon, for locally engaged staff to occupy key officer positions. There are, therefore, five cases in our data where the number of key offi-

cers reported at a mission exceeds the number of U.S.-based staff there. Reassuringly, we find no cases where the number of distinct key officers exceeds a mission's total staffing.

## Technical Validation 3: The Underlying Accuracy of the Key Officers Publications

Using the same OIG reports, we are also able to carry out a validation of the underlying accuracy of the *Key Officers* publications. Inspection reports routinely include a table of the "principal officials" of the inspected mission. The principal officers identified in the Historian database are Washington-based leaders at State Department headquarters. Because inspection reports are carefully fact-checked documents with potential career consequences for the named officials, we believe it is reasonable to treat these tables as the ground truth.

"Principal officials" are not precisely the same as "key officers"; the former category includes only individuals occupying leadership roles within the mission and is, thus, typically much smaller. Specifically, the OIG lists the chief of mission, deputy chief of mission, the chiefs of foreign service sections, the lead representative of other agencies, and leadership of any subordinate consular or other posts. Thus, not all "key officers" ought to have a matching listing within the OIG listing of "principal officials." We can, however, use the comparison for two purposes. First, when a given position is present in both lists, we can use this to assess the accuracy of the names listed in key officers (i.e., do the two sources identify the same position holder?). Second, we can use the share of "principal" positions listed in "key officers" to assess its comprehensiveness—that is, what percentage of leadership positions at diplomatic missions are represented in key officers?

We identify a total of 292 inspection reports of diplomatic missions from 2002 to 2017 with relevant data, covering a total of 4,369 principal officials at those missions. We begin by analyzing a subset of these positions where we are confident that *Key Officers* and

the "principal officials" listed in inspection reports should comprehensively overlap—the five core foreign service sections (consular, economic, management, political, and public affairs) at each post. There are 931 such positions reported in the OIG reports; of these, 897 (96.3%) correctly match the corresponding entry in *Key Officers*, meaning the same officer is listed in either the quarter of the inspection or the quarter immediately preceding it. There are 25 cases (2.6%) involving non-matches due to timing errors—that is, the officer from the OIG report is listed at *some* point in *Key Officers* but not in the correct quarter—and 8 cases (0.8%) of outright clashes where the officer in the OIG report never appears in the relevant position in *Key Officers*. We find only a single case where the entry in the OIG report has no corresponding entry in *Key Officers*. In sum, *Key Officers* is typically highly accurate with respect to the core foreign service positions, and approaches perfection if one makes some allowance for imprecise timing.

Turning now to the full sample of 4,369 principal officials, 73% match straightforwardly to *Key Officers* via an automated process. Of the 1,191 entries without an automated match, we randomly sample 200 for analysis. Within this sample, we manually find matches for 22 further cases not matched on automated basis, implying an overall match rate of about 76%. We find 34 mismatches, where OIG and *Key Officers* identify different occupants for the same position (17% of the unmatched subset; about 5% of the total) of which 25 are attributable to timing errors and 9 cases involve a complete clash (where the officer identified in the OIG report never appears in the relevant *Key Officers* position.)

In the substantial majority of the sample of unmatched cases (133 of 200, 67%), we find no listing in *Key Officers* corresponding to the position in the inspection report. That is, *Key Officers* frequently leaves out positions identified as "principal officials" by the inspection reports. Extrapolated to the full sample, this affects about 18% of reported principal officials. Notably, given that this happened only a single time for the foreign service sections, these omissions are almost exclusively related to non-State Department attachés at diplo-

matic missions. On a bureaucratic level, this is understandable: given that *Key Officers* is a State Department publication, it likely focuses more on State Department personnel.

Some omissions are systematic—that is, a given position is not reported at *any* post in a given quarter-year. For example, *Key Officers* reports Peace Corps positions from 1991 to 2003, but no post reports a Peace Corps position after 2003. Inspection reports show that many missions still had a Peace Corps presence after that date (75 different inspections list such an officer), but the State Department chose to stop listing these in *Key Officers*. On the other hand, some omissions are non-systematic—they affect some posts and not others. For example, *Key Officers* lists attachés from the Centers for Disease Control (CDC) at some diplomatic posts from 1990 onward, but other CDC attachés are omitted for no apparent reason. Inspection reports confirm, for example, that the CDC had attachés in Beijing and Copenhagen in 2010. *Key Officers* lists CDC attachés at forty diplomatic missions in that year, but omits these cities (among others). Eventually, in 2013, *Key Officers* begins to list the CDC position in Beijing. The position in Copenhagen is never listed.

Another set of omissions from *Key Officers* are evidently intended to shield the identity of CIA operatives. In 40 cases, inspection reports list a section chief for a "Regional Affairs" office—an amorphous title commonly used as cover for CIA stations [50, p. 26]. None of these have corresponding entries in *Key Officers*, and many officers who have held these positions can be publicly linked to the CIA. Likewise, we have found a number of cases where positions anomalously disappear from *Key Officers* and contemporaneous inspection reports show that the position still existed but was held by a CIA officer. For example, no political/military officer is listed in Key Officers for Abu Dhabi in 2010, though such a position is listed both before and after that time. A 2010 inspection report shows such a position held by Peter Enzminger, subsequently identified by journalists as a senior CIA officer [51].

Given that timing errors represent most of the outright clashes between the datasets,

Table 6: Quarter of Actual Arrival vs. Quarter First Reported in Key Officers

| First Appearance in Key Officers | Number | Percentage |
|---|---|---|
| Prior to Actual Arrival | 68 | 2% |
| Quarter of Actual Arrival | 273 | 8% |
| One Quarter After Actual Arrival | 1966 | 55% |
| Two Quarters After Actual Arrival | 548 | 15% |
| Three Quarters After Actual Arrival | 315 | 9% |
| One Year or More After Actual Arrival | 401 | 11% |

we now return to another way of measuring the accuracy of timing. The tables present in inspection reports typically indicate when each principal official arrived at the post in question. We can compare this to an officer's first appearance in *Key Officers* for that same role. Table 6 provides information on the difference between an officer's actual arrival at post (as reported in the OIG reports) and their first reported appearance at that post in Key Officers.

The modal officer appears in *Key Officers* the quarter after their arrival, and it is rare for officers to appear during or prior to their quarter of actual arrival. Given this, for certain purposes, it might be reasonable to offset the dataset by one quarter. There is, however, fairly wide variation and fully a tenth of officers do not appear in *Key Officers* until a year after their actual arrival. Overall, this suggests that the specific timing reported in KOFSP should be treated with caution—the data cannot identify with precision exactly when officers arrived or departed. On the other hand, the names given in KOFSP are nearly always accurate aside from the imprecise timing information.

# Usage Notes

We conclude by discussing practical considerations of concern to any potential user of the KOFSP data.

Table 7: Career coverage: Ann B. Sides

| Position | Years | In KOFSP | Why not? |
|---|---|---|---|
| Consular Officer, Niamey | 1984-1985 | No | Too junior |
| Consular Officer, Oran | 1985-1987 | Yes | – |
| Consular Officer, Dakar | 1987-1990 | Yes | – |
| Consular Officer, Belgrade | 1991-1992 | Yes | – |
| Consular Officer, Dublin | 1993-1996 | Yes | – |
| West African Affairs Desk, DC | 1996-1998 | No | In DC |
| Consular Officer, Sarajevo | 1998-2001 | Yes | – |
| Consul General, Belgrade | 2001-2004 | Yes | – |
| Consul General, Athens | 2004-2008 | Yes | – |
| Grievance Officer, DC | 2008-2010 | No | In DC |

## Positions covered vs. not covered in KOFSP

First, to reiterate, the KOFSP data is *not* a comprehensive list of all foreign service officers, nor is it a comprehensive list of all individuals serving at U.S. diplomatic posts. Further, it is not a random sample of these individuals; officers deemed "key" are typically the highest-ranking in their position and post. Thus, the data should be seen primarily as a record of the individuals holding important positions at U.S. diplomatic posts over time, and not as an attempt to capture the entire U.S. diplomatic workforce. Relatedly, because officers may hold a combination of "key" and non-"key" positions over the course of their careers—including, for most foreign service officers, at least one domestic posting in Washington D.C.—the diplomat-level career trajectories in our data are not exhaustive lists of all positions an officer has held. Rather, they are records of all of an officer's appearances in the *Key Officers* directories over time.

To drive home this point, in Tables 7 and 8 we draw on ADST oral history interviews to compile complete career histories for two former diplomats—Ann B. Sides and Amb. Lange Schermerhorn—and compare these to our records of these officers in the KOFSP data. Sides, a consular-coned foreign service officer, spent most of her career as the highest-ranking Consular Officer at relatively small posts before serving as Consul General in Belgrade and Athens. As a result, she appears in the KOFSP data for most of

Table 8: Career coverage: Amb. Lange Schermerhorn

| Position | Years | In KOFSP | Why not? |
|---|---|---|---|
| Consular Officer, Colombo | 1967-1969 | Yes | – |
| Consular Officer, Saigon | 1969-1970 | No | Too junior |
| DOS Operations Center, DC | 1971-1973 | No | In DC |
| Economic Officer, Tehran | 1974-1978 | No | Too junior |
| Commercial Officer, London | 1978-1981 | No | Too junior |
| National War College, DC | 1981-1982 | No | In DC |
| North African Affairs Desk, DC | 1982-1985 | No | In DC |
| Economic Officer, Brussels | 1985-1988 | Yes | – |
| DOS Economic Bureau, DC | 1988-1990 | No | In DC |
| DOS Office of Career Development, DC | 1990-1993 | No | In DC |
| Deputy Chief of Mission, Brussels | 1993-1997 | Yes | – |
| Ambassador to Djibouti | 1998-2001 | Yes | – |

her career. Still, however, we do not capture her first consular tour (she was too junior), nor do we capture her two postings in Washington.

Ambassador Schermerhorn, on the other hand, held a number of different positions at relatively large and high-importance posts such as U.S. Embassy London and U.S. Embassy Tehran early on in her career; she was not listed in the *Key Officers* directories during these postings, as she would not have been the top-ranking officer in her section at these posts. Further, she spent an unusually large amount of time in Washington during her career; as such, the KOFSP data do not record the bulk of her mid-career positions. As a result, despite a long diplomatic career which ultimately resulted in an ambassadorial appointment, Amb. Schermerhorn only appeared in the *Key Officers* directories in four of her seven overseas postings (and twelve total career postings).

## Changing coverage over time

As noted above, the positions reported as "key" in the directories change over time. There is both systematic variation in reporting at the quarter-year level (i.e., certain positions are reported only in certain quarter-years) and idiosyncratic post-level variation (i.e., certain posts either idiosyncratically report or fail to report particular positions). The

Table 9: Coverage of the Key Officers Data Over Time

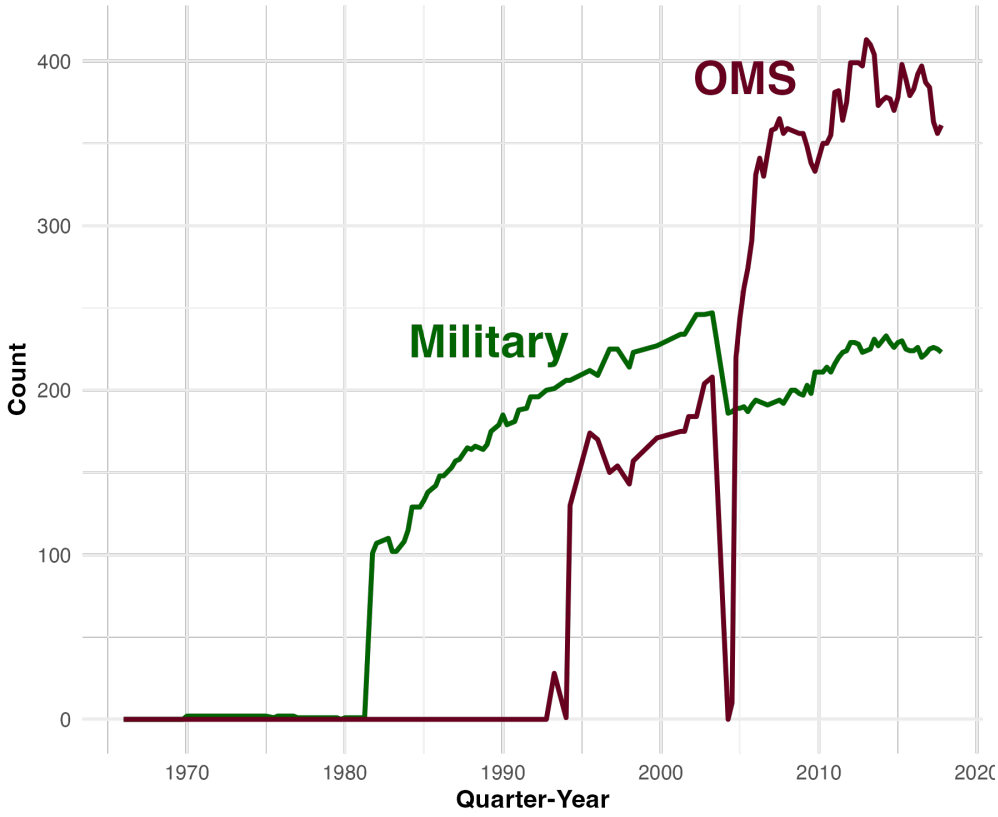
| Year | Generalist Key Officers | Specialist Key Officers | Total FS Key Officers | Total FS | % of FS in KOFSP |
|---|---|---|---|---|---|
| 1970 | 1,070 | 0 | 1,070 | 8,932 | 12.0% |
| 1980 | 1,132 | 66 | 1,198 | 9,326 | 12.8% |
| 1985 | 1,199 | 82 | 1,281 | 9,441 | 13.6% |
| 1990 | 1,227 | 140 | 1,367 | 8,814 | 15.5% |
| 1995 | 1,287 | 585 | 1,871 | 8,207 | 22.8% |
| 2000 | 1,273 | 601 | 1,872 | 9,238 | 20.3% |
| 2005 | 1,537 | 1,210 | 3,029 | 11,238 | 27.0% |
| 2010 | 1,586 | 1,587 | 3,484 | 12,859 | 27.1% |
| 2015 | 1,686 | 1,711 | 3,756 | 13,872 | 27.1% |

general, but not universal, trend here is towards reporting a greater number of positions over time. This is especially true for foreign service specialists and external attachés. Table 9 documents the basic phenomenon over time. The table shows the number of generalist, specialist, and total foreign service personnel reported in KOFSP over time versus the size of the overall foreign service. Note that the total column is not the sum of the generalist and specialist columns because it also includes unclassified foreign service officers.

At the beginning of the KOFSP data, the key positions reported cover 12% of the total foreign service. By 2015, this has more than doubled to 27.1%, as the set of positions reported as "key" has expanded. This increase in the portion of the foreign service covered by KOFSP is almost entirely driven by specialists: between 1970 and 2015, the foreign service grew by 55%, while the number of generalists reported in KOFSP grew by almost the exact same amount, 58%. On the other hand, *no* specialists are reported in KOFSP in 1970; by 2015, more specialists than generalists are reported.

For many positions, it is possible to identify a discrete moment in time when coverage changes for a particular role, presumably reflecting some central decision to either designate or de-designate a given position as a "key" one for reporting followed by rapid diffusion. Figure 4 highlights two illustrative examples from the data, showing changes in the number of military attachés and office management specialists (OMS) reported in

Figure 4: Reporting of select positions: Military Attaché and OMS




KOFSP over time.

Military attachés have been present at US embassies since the nineteenth century; however, KOFSP does not initially capture them. Until 1981, the only military attachés reported are the defense advisor to U.S. Mission to NATO and (in some quarters) their deputy. In the third quarter of 1981, *Key Officers* suddenly begins to report military attachés at many posts, leading to an era of steady growth until another discontinuity in 2004, when the number of reported attachés drops from 223 to 174. Rather than representing a decline in the military presence at embassies, this reflects an apparent decision in 2004 to stop reporting certain military positions. In 2003, *Key Officers* includes both the heads of Defense Attaché Offices (DAO) and the heads of Offices of Defense Cooperation (ODC). In 2004, *Key Officers* abruptly stops reporting ODC positions.

As a second example, Figure 4 also plots the number of officers listed in "office man-

agement specialist" roles (i.e., administrative assistants to senior diplomats). No such position is reported in the data until 1993, when a few instances appear, only to disappear in the next quarter. In the second quarter of 1994, reporting of OMS roles becomes widespread and remains so for a decade with somewhere between 150 and 200 listings. Nearly all of these listings anomalously disappear in 2004 before reappearing; after this, the number reported rapidly reaches over 300. This second surge appears to largely relate to the increased listing of DCM OMS position (i.e, an OMS assisting the Deputy Chief of Mission)—22 of these are listed in 2003, compared to 126 in 2006.

As is evident from these examples, users of the data should exercise caution in drawing inferences from the absence of a listed position at a given post in a given time period. The absence of a listing is clearly uninformative when no such position is listed anywhere (e.g., the absence of a military attaché listing before 1981 at a post is meaningless). In cases where the position is listed at *some* posts, the absence of a listing at a particular post may be meaningful, but we suggest that users interested in such questions examine the relevant data carefully and consult external documentation when available. We note that the data are least reliable for external attachés. In many cases, these officials can alternatively be located through publicly available personnel records associated with their parent agencies and users focused on attachés may wish to consider this option.

The change in coverage over time also has important implications for analyzing longitudinal changes at a given post. An apparent increase in the number of officers at a post may reflect either actual growth or simply increased reporting. Users interested in changes over time should consider benchmarking the post of interest against other posts to gauge whether changes likely come from differential reporting or actual changes in size. Again, special care should be taken with respect to attachés and inferences based on absence. For example, we do not believe that KOFSP can be reliably used to answer questions such as when a given post first received an attaché of a particular type.

We note that the core foreign service generalist positions at embassies and consulates

general are consistently reported over time with no apparent changes in coverage. We consider these the most reliable subset of the data.

## Residence

In the Creating Other Variables section, we explain our coding process for three variables pertaining to the officer's physical location of residence (`residence_location`, `residence_certainty`, and `officer_is_at_post`). Here we highlight a number of practical considerations for users of the data to keep in mind regarding questions of officer residence.

First, we note that officer residence locations are most uncertain among external attachés: among position categories "External" and "Nonstate FSO", 20% of observations are coded `officer_is_at_post` = 0, while the corresponding figure for all other position categories is 4%. External attachés are especially overrepresented among officers with many simultaneous listings. There are 339 cases of officers listed at more than ten posts in the same quarter; 296 of these are External officers, and 36 are Generalists in a "Labor" position. Examining these outliers, we find some notable inconsistencies in reporting practices. Kathy Beck (officer ID: 18658), for example, is listed as IRS attaché at between 46 and 48 different posts across Europe, Africa, and the Middle East in each quarter of 2007–2008, while resident in Paris. In contrast, consider Jim Coffey (officer ID: 12026), FAA attaché posted in Frankfurt in the 1990s. From 1991–1994q1, he is listed exclusively at Frankfurt; in 1994q2, he is listed at 52 other posts; but for the five subsequent quarters of his posting, he is listed at no more than six other posts.

How these cases of non-resident attachés should be handled depends on the purpose of the analysis. If a researcher is interested in studying an individual officer's career trajectory, it would likely make sense to only focus on postings where an officer is physically resident. Alternatively, if the goal is to understand whether a particular agency's services or expertise is being provided at a given post, it is unclear how to best proceed. It seems

unlikely that Jim Coffey was actively engaging with all 53 posts listing him simultaneously in 1994q2. It also seems unlikely that the extent to which he provided those posts with aviation policy expertise meaningfully changed between 1994q1 and 1994q2. Users may consider various approaches such as inversely weighting an attaché'ss contribution to a given post by the number of other posts they are serving simultaneously, and interpolating attaché presence at a given post over time.

## Data Availability

All four KOFSP datasets, a detailed codebook, and an archive of all 156 Key Officers source documents can be downloaded directly from our Harvard Dataverse site (https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/3IPF8Y) [48]. While the Dataverse site will only ever contain the version of the data that existed at time of publication, we plan to continuously release new iterations of the data as we identify and correct errors, add additional directories to the sample, and/or create additional variables. *Key Officers* documents continue to be published, creating the possibility of updating the temporal coverage of the data. Further, while no AI products were used to compile the original KOFSP datasets, we believe that recent advances in AI-based optical character recognition (OCR) and record-linking techniques could dramatically reduce the time required to update the data. Readers interested in checking for updates can do so by consulting https://measuringdiplomacy.github.io.

## Code Availability

Code and data necessary to (1) assemble each of the aggregate datasets from the officer-position-quarter-year dataset, and (2) reproduce all figures in the paper can be downloaded at https://github.com/measuringdiplomacy/kofsp_replication_files.

## Acknowledgments

The authors wish to acknowledge Nate Jensen, who supported this project in its earliest stage; HathiTrust and the Internet Archive, organizations without which this data collection effort would not have been possible; and the many current and former foreign service officers who shared their insights with us as we sought to classify and make sense of what we had collected.

## Author contributions

M.M. and C.T. collected, digitized, and transcribed the *Key Officers* documents. D.L., M.M., and C.T. all contributed equally to (1) cleaning the data; (2) record-linking of individual diplomats; (3) producing created variables from the data; (4) writing this article, including generating tables and figures.

## Competing interests

The authors declare no competing interests.

## Funding

The authors received no external funding.

# References


[1] David Lindsey. Diplomacy Through Agents. *International Studies Quarterly*, 61(3):544–556, September 2017.

[2] David Lindsey. *Delegated Diplomacy: How Ambassadors Establish Trust in International Relations*. Columbia University Press, New York, 2023.

[3] Matt Malis and Calvin Thrall. Diplomatic capacity and international cooperation. Unpublished Manuscript, 2025.

[4] Minju Kim and Shu Fu. Bringing Home the Bacon: Politician Ambassadors and Home State Trade. *World Politics*, 77:382–417, 2025.

[5] Eric Lipton, Nicola Clark, and Andrew W. Lehren. Diplomats help push sales of jetliners on the global market. *The New York Times*, January 2011.

[6] Selwyn J. V. Moons and Peter A. G. van Bergeijk. Does Economic Diplomacy Work? A Meta-analysis of Its Impact on Trade and Investment. *The World Economy*, 40(2):336–368, February 2017.

[7] Jakob Munch and Georg Schaur. The Effect of Export Promotion on Firm-Level Performance. *American Economic Journal: Economic Policy*, 10(1):357–387, February 2018.

[8] Calvin Thrall. Informational lobbying and commercial diplomacy. *American Journal of Political Science*, 69(3):1147–1162, 2025.

[9] Geoffrey Gertz. Commercial diplomacy and political risk. *International Studies Quarterly*, 62(1):94–107, 2018.

[10] Julia Gray and Philip Potter. Diplomacy and the Settlement of International Trade Disputes. *Journal of Conflict Resolution*, 64(7-8):1358–1389, August 2020.

[11] Matt Malis. Conflict , Cooperation , and Delegated Diplomacy. *International Organization*, 2021.

[12] Alan Greenblatt. Diplomatic cables: The ties that bind foreign policy. *NPR*, Dec 2010.

[13] Mary Thompson-Jones. *To the Secretary: Leaked Embassy Cables and America's Foreign Policy Disconnect*. W.W. Norton, New York, 2016.

[14] Matt Malis. Status and Influence in the Foreign Policy Bureaucracy. Working Paper, 2023. https://mattmalis.github.io/files/pdf/Malis_Influence_31mar2023.pdf.

[15] Romain Lecler and Yann Goltrant. Gendering diplomatic careers. distance and time in international assignment practices among 600 french diplomats. *Gender, Work & Organization*, 32(5):1944–1965.

[16] David Lindsey. Who decides who gets in? diplomats, bureaucrats, and visa issuance. *Political Science Research and Methods*, 13(4):850–864, 2025.

[17] Evan T Haglund. Striped pants versus fat cats: Ambassadorial performance of career diplomats and political appointees. *Presidential Studies Quarterly*, 45(4):653–678, 2015.

[18] Gary E Hollibaugh. The political determinants of ambassadorial appointments. *Presidential Studies Quarterly*, 45(3):445–466, 2015.

[19] Eric Arias and Alastair Smith. Tenure, promotion and performance: The career path of US ambassadors. *Review of International Organizations*, 13(1), 2018.

[20] Michael A Goldfien. Just patronage? familiarity and the diplomatic value of non-career ambassadors. *Journal of Conflict Resolution*, 68(7-8):1417–1442, 2024.

[21] Birgitta Niklasson and Ann Towns. Gender, status, and ambassador appointments to militarized and violent countries. In *Gender and Diplomacy*. Routledge, New York, 2017.

[22] Birgitta Niklasson and Ann E Towns. Diplomatic Gender Patterns and Symbolic Status Signaling: Introducing the GenDip Dataset on Gender and Diplomatic Representation. *International Studies Quarterly*, 67(4):sqad089, 11 2023.

[23] Ryan M Scoville. Unqualified ambassadors. *Duke Law Journal*, 69(71):73–196, 2019.

[24] Sabrina B. Arias. Who sets the agenda? diplomatic experience and small power influence in the united nations. *World Politics*, 78(3), 2026.

[25] Reşat Bayer. Diplomatic Exchange Data set, v2006.1., 2006.

[26] Jonathan D Moyer, Sara D Turner, and Collin J Meisel. What are the drivers of diplomacy? introducing and testing new annual dyadic data measuring diplomatic exchange. *Journal of Peace Research*, 58(6):pp. 1300–1310, 2021.

[27] Tyler Jost and Yucong Li. The politics of promotion in china's foreign policy bureaucracy. *The China Quarterly*, 263:776–795, 2025.

[28] Andrew K. Rose. The foreign service and foreign trade: Embassies as export promotion. *World Economy*, 30(1):22–38, 2007. ISBN: 1007029395222.

[29] Lauge N.Skovgaard Poulsen and Emma Aisbett. Diplomats want treaties: Diplomatic agendas and perks in the investment regime. *Journal of International Dispute Settlement*, 7(1):72–91, 2016.

[30] Jonathan Renshon. Status deficits and war. *International Organization*, 70(3):513–550, 2016.

[31] Melvin Small and J David Singer. The diplomatic importance of states, 1816–1970: An extension and refinement of the indicator. *World Politics*, 25(4):577–599, 1973.

[32] Benjamin O Fordham. Who wants to be a major power? explaining the expansion of foreign policy ambition. *Journal of Peace Research*, 48(5):587–603, 2011.

[33] Brandon J Kinne. Dependent diplomacy: Signaling, strategy, and prestige in the diplomatic network. *International Studies Quarterly*, 58(2):247–259, 2014.

[34] Marina G Duque. Recognizing international status: A relational approach. *International Studies Quarterly*, 62(3):577–592, 2018.

[35] Paul Sharp. For diplomacy: Representation and the study of international relations. *International studies review*, 1(1):33–57, 1999.

[36] Emanuele Colonnelli, Mounu Prem, and Edoardo Teso. Patronage and selection in public sector organizations. *American Economic Review*, 110(10):3071–99, October 2020.

[37] Jan H Pierskalla, Adam Lauretig, Andrew S Rosenberg, and Audrey Sacks. Democratization and representative bureaucracy: An analysis of promotion patterns in indonesia's civil service, 1980–2015. *American Journal of Political Science*, 65(2):261–277, 2021.

[38] Joachim Voth and Guo Xu. Discretion and destruction: Promotions, performance, and patronage in the Royal Navy. Technical report, CEPR Discussion Paper No. DP13963, 2022.

[39] Didac Queralt. Her majesty's aid: Imperial subsidies and state building in the late british empire. 2025. https://www.didacqueralt.com/_files/ugd/8b1fc6_2602c8ba328a4041b854f61b9e232aef.pdf.

[40] Juan Felipe Riano. Bureaucratic nepotism. *Working Paper*, 2024.

[41] Matthew Weed and Nina Serafino. *U.S. Diplomatic Missions: Background and Issues on Chief of Mission (COM) Authority*. Congressional Research Service, Washington, DC, 2014.

[42] John Marks. How to spot a spook. *The Washington Monthly*, pages 5–11, November 1974.

[43] U.s. halts two booklets used to spot c.i.a. men. *The New York Times*, mar 1976.

[44] Department of State. Key Officers Booklet: New Embassy CERP Requirement (Airgram 3743). *State Department Central Files (NARA P-Reels)*, P780163-1303, 1978.

[45] U.S. State Department. Commercial program: Updating of guide for business representatives for senior commercial. Cable, 1977. Addressed to all diplomatic and consular posts.

[46] Harry W. Kopp and John K. Naland. *Career Diplomacy*. Georgetown University Press, 2017.

[47] Vincent Arel-Bundock, Nils Enevoldsen, and Cj Yetman. countrycode: An r package to convert country names and country codes. *Journal of Open Source Software*, 3(28):848, 2018.

[48] David Lindsey, Matt Malis, and Calvin Thrall. Key officers of u.s. foreign service posts (kofsp) data, 2025. https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/3IPF8Y.

[49] Office of the Historian, Foreign Service Institute, United States Department of State. Principal officers and chiefs of mission, 2024.

[50] Bryan Denson. *The Spy's Son: The True Story of the Highest Ranking CIA Officer Ever Convicted of Espionage and the Son He Trained To Spy For Russia*. Atlantic Monthly Press, New York, 2015.

[51] Tim Weiner. *The Mission: The CIA in the 21st Century*. Mariner Books, New York, 2015.