



UNIVERSITY OF GOTHENBURG

GenDip

Dataset on Gender and Diplomatic Representation, 1968 - 2024

Version: June 2025

Authors: Ann E. Towns and Birgitta Niklasson

Citation

Scholars who wish to use this dataset in their research are kindly requested to use the following citation:

Towns, Ann E. and Birgitta Niklasson (2025) *The GenDip Dataset on Gender and Diplomatic Representation, version June25*. University of Gothenburg: The GenDip Program: www.gu.se/en/gendip/gendip-data

Acknowledgements

The June25 version of the GenDip Dataset would not have been possible without the leadership and hard work of GenDip program coordinator Kajsa Evertsson and Natalia Natsika. Kajsa was responsible for supervising the interns who did the coding of the 2024 postings during fall of 2024 and spring of 2025, including quality checking their work. Kajsa has also identified missing missions and postings in the previous versions (1968-2021) which has enabled us to provide more complete data. She also coded some difficult cases, missing and incomplete data, as well as resolved uncertain coding cases for 2024. Natalia Natsika brought her extensive experience of working with datasets from V-Dem to GenDip. In addition to double checking all the manually coded data 2014-2021 against the Europa World Yearbook (EWY), Natalia has also helped systematize the data presentation and align it with standards in the field, and clean the data. Finally, Natalia has coded some difficult cases, missing and incomplete data, as well as resolved uncertain coding cases for 2024.

Interns, research assistants, colleagues, and friends who have coded data or helped out with the dataset in other ways (alphabetically): Naomi Agabu, Joseph Akowuah, Sara Sofia Anderson, Ingrid Andreasson, Ylva Beckman, Rebecca Billeci, Ketevan Bolkvadze, Arash Bolouri, Kristin Boxström Le Guillarme, Lars Crusefalk, Ani Dimitrova, Beatrice Engman, Linus Forsman, Sindre Gade Viksand, Moosa Ghaeb, Kristina Gurok, Mais Haddad, Rabab Karim Hashish, Felicitas Höck, Katarzyna Jezierska, Sahithi Kanapala, Chuleeporn Karlsson, Yunie Kim, Farman Ali Khan, Anne- Kathrine Kreft, Sofia Laquaglia Galliano, Karolina Lecalonnec, Rafael Lopez Valverde, Haley Mcewen, Thu Lê, Mino Nakamura, Reachnithor Oum, Jasmin Rath, Felicia Robertson, Kristi Rød Gulliksen, Tove Selnes, Hasna Shawaly, Savina Sirik, Victoria Sjöberg, Signe Stahlenius, Tove Wikehult, Tsz-wa Yip, Kaisa Yli-Rohdainen, Lixia Zheng.

Colleagues who have provided insightful and helpful comments (alphabetically): Stephen Dawson, Monika de Silva, Katarzyna Jezierska, Andrej Kokkonen, Elias Markstedt, Marina Nistotskaya.

Table of contents

The GenDip Program	4
Contents of the Dataset	4
Aim of the Dataset	4
The Data Collection Process	5
Variables	6
<i>year</i>	6
<i>cname_send</i>	6
<i>Main_posting</i>	6
<i>title</i>	7
<i>gender</i>	8
<i>cname_receive</i>	8
<i>ccode_send</i>	9
<i>ccodealp_send</i>	9
<i>ccodeCOW_send</i>	9
<i>region_send</i>	9
<i>ccode_receive</i>	10
<i>ccodealp_receive</i>	10
<i>ccodeCOW_receive</i>	10
<i>region_receive</i>	10

The GenDip Program

The main aim of the GenDip program is to identify and analyze the changing nature of gender in and of diplomacy. The program constitutes an integrated part of the Department of Political Science at the University of Gothenburg and was founded by Professor Ann Towns in 2014.

Contents of the Dataset

The GenDip dataset includes the gender classification, diplomatic title, sending state and receiving state of all bilateral ambassadors heading embassies in the world 1968–2024, **89,446 ambassador postings in total**. The dataset is structured as dyad/decade for the period 1968–1998 and dyad/lustrum for 1998–2024. It starts in 1968 and then relies on 10-year increments until 1998 because of the consistently low share of women during the 1968–1998 period. In all, the dataset spans the time from when female heads of mission constituted rare exceptions until 2024, when women made up **23,4%** of all bilateral ambassadors and headed a majority of some states' foreign missions.

Aim of the Dataset

Diplomatic missions are an important link between actors in international relations. As representatives of sending states to receiving states, diplomats serve as important channels of communication and information between states (but also between other actors). The establishment of a diplomatic mission in another state or territory thus signals recognition and that the state is worthy of official communication (e.g. Singer and Small 1969; Small and Singer 1973; Li 2013; Kinne 2014).

In addition to *where* a mission is placed, there are of course many other significant aspects of a diplomatic mission, including traits of the head of mission. For instance, previous research has pointed to the importance of the diplomatic title of the head of mission (Moyer et al. 2020). One crucial dimension in diplomatic representation has been overlooked, however: the gender of the head of mission.

In the nineteenth and twentieth centuries, most diplomats were men. But in the twenty-first century, the share of female diplomats has increased markedly. When and where have the shares of female ambassadors increased most? What, if anything, happens to diplomacy when women come onto the stage as diplomatic officials? How does this shape relations between states? Why do some states send so much higher shares of male ambassadors than others? What might the posting of a male vs a female ambassador signal to the receiving government? To home audiences? Given the symbolic importance of the status of women for how states promote and perceive themselves (e.g. Towns 2010), answering questions about which states appoint women (or men) as heads of mission and to where is likely to provide new and exciting insights about states' pursuit of status. The data can also be used for a range of other intriguing questions about international politics.

The purpose of this dataset is to provide information on the gender of diplomats around the world, data that has so far been missing from existing datasets on diplomatic exchanges (Bayer 2006; Rhamey et al. 2013; Moyer et al. 2020). Our hope is thus that the GenDip dataset will supply the fields of international politics and gender and politics – including the exciting and fast-growing research field on the gender of diplomacy – with a foundation to be explored and developed in a range of ways.

The Data Collection Process

The *Europa World Year Book (EWY)* series is the standard source for diplomacy datasets. *EWY* provides the name, sending country and receiving country of ambassadors but does not provide information on their gender classification. How *EWY* collects its information and its accuracy is not entirely clear. As a result, our dataset for the years 2014 onward does not only include information from *EWY* but other sources as well. See below for a clarification of the data collection process for the years coded. We have been very careful in the data collection process and in cases where we have not been entirely sure about the coding, we have put in a missing value rather than a value that could potentially be incorrect.

1968-2013: Since it is progressively more difficult to find diplomatic lists and MFA publications back in time, we opted to rely only on information from *EWY* for 1968–2013. Since *EWY* includes information about ambassadors' names and not gender, the gender coding is almost exclusively based on the diplomats' names. That coding necessitated involving and hiring coders with cultural name recognition competence for different parts of the world. See more information below under the variable description for gender.

2014-2021: Data was collected from three sources: (1) information from sending states, e.g., the websites of the sending MFAs, official MFA yearbooks, biographical ambassador information (e.g., CVs), and email inquiries to the MFAs; (2) diplomatic lists, which identify all foreign diplomats in a host state, produced by the host state MFA; and (3) data from *EWY*. Data from *EWY* was used as cross-check of the data collected from (1) and (2), and not as the primary data collection source. After cross-checking our own data with *EWY*, we found a few discrepancies, primarily postings that the GenDip interns had missed due to lack of reliable online sources. In those few instances, we were able to find reliable information (typically from the sending MFA) to provide a corrected coding.

2024: The data was collected from two sources: (1) information from sending states, e.g., the websites of the sending MFAs, official MFA yearbooks, biographical ambassador information (e.g., CVs), and email inquiries to the MFAs; and (2) diplomatic lists, which identify all foreign diplomats in a host state, produced by the host state MFA. The data derives primarily from (1) information from sending states and has been cross-checked against (2) diplomatic lists for some cases but not for all of them. The data collection took place during autumn 2024 and spring 2025. As a general rule, we have coded the last appointed ambassadors and not the ambassadors who held the post the longest within the year. This means that if an ambassador changed during 2024, we have coded the ambassador in charge during autumn 2024 and not the ambassador that was in charge before autumn 2024.

Additionally, we will produce a new updated version of the dataset in August 2025 and include the third cross-check; (3) data from *EWY* for 2024. *EWY* data from 2024 was not released until mid-June 2025 and we therefore have not had the possibility to include the cross-check in this June25 update. We expect very little to change as a result of this check.

Variables

year

The year for which the data is valid. The data collection for the *Europa World Yearbook (EWY)* takes place the year before the *EWY* is published, which for example means that the year 1968 in the dataset is taken from the *EWY* published in 1969.

cname_send

Name of the sending country. The name is written according to the ISO 3166-1 standard: <https://www.iso.org/obp/ui/#search>

Main_posting

The variable *main_posting* refers to the main (and resident) posting of the diplomat.

Ambassadors are often accredited to several countries simultaneously, so that one individual serves as main and resident ambassador in one capital and as a concurrently side-accredited ambassador to others. We thus created a variable, *main_posting*, to indicate what is usually the resident posting. This variable enables analyses of individuals rather than postings (as one individual can have several postings - though only one main posting - per year).

As a general rule, we have tried to code the last appointed ambassador for each year update, and not the ambassador who held the posting the longest within a given year. For example, if an ambassador changed during a year, we have tried to code the ambassador in charge during autumn of that year. However, there are cases where this has not been possible throughout, and there will be some discrepancies in this regard, either due to lack of updated information for the whole year, or due to coding having taken place in earlier stages within the year.

Because of the difficulties of finding complete information on side accreditations, side postings are not included in the dataset.

Occasionally, a diplomat may not live in the country of their main posting. Typically, in such cases, the diplomat either resides in the home capital or (temporarily, such as during a war) in another country. For the 2024 data, we have removed all non-resident main postings we were able to identify. However, it is possible that a few non-resident main postings have nonetheless been included for 1968-2021 because the information was unclear. In our assessment, this is a very small number of postings.

Any given year, each diplomat is coded with one main posting. However, in about 800 cases during 1968-2009, the *EWY* does not clearly specify which of several posts is the main one. We have then coded as “main_posting” the posting of: 1) the highest rank (e.g. if a diplomat is an ambassador in one country and “only” chargé d'affaires in another, and 2) the country with the highest GDP the relevant year (World Bank website).

Sometimes, states close embassies. If we can confirm that an embassy has been closed, we indicate this with the code 98.

Sometimes, when a state is in serious turmoil, the status of its sent and received missions is unclear. Posts will be vacant, embassy operations temporarily suspended, and diplomatic staff may be relocated to the home capital or another country. Afghanistan in 2021 and 2024 is such a case. All Afghanistan's potentially received and sent heads of mission in 2021 and 2024 were thus coded as “missing data” (99), unless we could confirm that the embassy had indeed been permanently closed (98).

Sometimes, the status of a polity is unclear, making the status of its mission and diplomatic representative unclear as well. Only recognized states may send and receive ambassadors. Palestine is one such case. Regarding Palestine as a receiving country, we have included ambassadors of two confirmed embassies in Palestine in the dataset, but we have not included heads of “permanent representative offices” to Palestine. Regarding Palestine as a sending country, we have included main postings sent from Palestine that have been coded by the GenDip interns, but we have not managed to conduct quality controls for this information against the *EWY* because it does not include sent postings from Palestine. We therefore acknowledge that we have handled Palestine in a different way in order not to lose valuable information on its missions sent abroad, but we aim to address this in the next data update.

1 = main posting

98 = embassy closed

99 = missing data

title

Title of the diplomat.

Heads of diplomatic missions can rank as ambassadors, chargé(e)s d'affaires, and more.

Ambassadors enjoy the highest rank, whereas the lower ranked chargé(e)s d'affaires may serve as the (often temporary) chief of mission in the absence of an ambassador.

1 = chargé d'affaires

2 = minister, internuncios

3 = ambassador, high commissioner, papal nuncios (e.g. pro-nuncios, apostolic nuncio, etc.), designate ambassador, ambassador-at-large, secretary of the people's bureau (Libya)

4 = other (e.g. acting ambassador, acting high commissioner, chargé d'affaires ad interim, envoy)

98 = embassy closed

99 = missing data

gender

The gender classification of the diplomat.

The coding is based on titles (e.g. Mr/Mrs, prince/princess, baron/baroness, etc.), pronouns used when referring to the diplomat (she/her, he/him, they/them, etc), or the recognition of names as either female or male. We have not yet encountered an ambassador that is classified as something other than man or woman.

Sometimes, the *EWY* only states the initials for the first names of mission heads. Thus, for 1968-2009, if the diplomats have not been known to us through other sources and their names do not reveal their gender classification, their gender has been coded as missing.

In all, the dataset includes gender classification codes for 93% of the total 89 446 postings. There is some variation across time; we were able to assign gender classification to a larger share of ambassadors for more recent postings than for postings in 1968 and 1978, partially because only the initials of the ambassadors' names were included in the *EWY* in those early years. There is also some variation across countries, with a lower share of gender classifications for appointments made by a few smaller Asian states.

Our data for 2014, 2019, 2021, and 2024 is a more reliable source for gender classification than *EWY* since we have been able to categorize all diplomats based on some combination of gendered title, gender pronoun, and name. For the gender coding 1968-2008 (when we relied only on the *EWY* books), the gender coding is almost exclusively based on the diplomats' names. That coding necessitated involving coders with cultural name recognition competence for different parts of the world. More specifically, we hired coders with cultural name recognition competence for China, Taiwan, Japan, Korea, Slavic language countries, Latin American countries, Sub-Saharan Africa, Egypt, Lebanon, Algeria, Liberia, Morocco, Iran, Pakistan, Bangladesh, Afghanistan, India, Vietnam, Thailand, Malaysia, Cambodia, Singapore, and Finland. Coders were instructed to be conservative and cautious – if there was any doubt about a binary gender classification of a name, they were instructed to code the gender as “missing” (99). Names that can refer to women as well as men have thus been coded as missing, unless more information about the ambassador has been available through other sources.

0 = man

1 = woman

98 = embassy closed

99 = missing data

cname_receive

Name of the receiving country. The name is written according to the ISO 3166-1 standard: <https://www.iso.org/obp/ui/#search>

ccode_send

Country numeric code of the sending country based on ISO 3166-1 numeric codes and ISO 3166-3. The latter standard specifies former codes for countries that do no longer exist. All numeric country codes are unique, which makes this variable the best suited one for merging datasets (in combination with year).

Source: <https://www.iso.org/obp/ui/#search> and <https://www.iso.org/standard/72484.html>

9999 = missing data

ccodealp_send

A three-letter code of the sending country based on the ISO-3166-1 alpha3 English standard.

Source: <https://www.iso.org/obp/ui/#search/code/> and <https://www.iso.org/standard/72484.html>

When information has been unavailable at the ISO page, we have also used information from Statistics Canada: <https://www.statcan.gc.ca/en/subjects/standard/sccai/2011/scountry-desc>

9999 = missing data

ccodeCOW_send

Correlates of Wars (COW) country code number according to Singer and Small (1994).

9999 = missing data

region_send

Geographical region of the sending country. *NB!* This coding does not take political or cultural affiliations into account.

0 = Africa

1= Asia

2 = Central and North America (including the West Indies)

3 = Europe (including Russia; excluding Nordic countries)

4 = Middle East (including Egypt, Israel, and Turkey)

5 = Nordic countries

6 = Oceania

7 = South America

ccode_receive

Country numeric code of the receiving country based on ISO 3166-1 numeric codes and ISO 3166-3. The latter standard specifies former codes for countries that do no longer exist. All numeric country codes are unique, which makes this variable the best suited one for merging datasets (in combination with year).

Source: <https://www.iso.org/obp/ui/#search> and <https://www.iso.org/standard/72484.html>

9999 = missing data

ccodealp_receive

A three-letter code of the sending country based on the ISO-3166-1 alpha3 English standard.

Source: <https://www.iso.org/obp/ui/#search/code/> and <https://www.iso.org/standard/72484.html>

When information has been unavailable at the ISO page, we have also used information from Statistics Canada: <https://www.statcan.gc.ca/en/subjects/standard/sccai/2011/scountry-desc>

9999 = missing data

ccodeCOW_receive

Correlates of Wars (COW) country code number according to Singer and Small (1994).

9999 = missing data

region_receive

Geographical region of the receiving country. *NB!* This coding does not take political or cultural affiliations into account.

0 = Africa

1 = Asia

2 = Central and North America (including the West Indies)

3 = Europe (including Russia; excluding Nordic countries)

4 = Middle East (including Egypt, Israel, and Turkey)

5 = Nordic countries

6 = Oceania

7 = South America