

# Comparing Climate Change Policy Networks: Codebook for Twitter Account Collection<sup>\*</sup>

Ted Hsuan Yun Chen<sup>1,2</sup>   Mikko Kivelä<sup>1</sup>   Arttu Malkamäki<sup>2</sup>  
Esa Palosaari<sup>1</sup>   Anniina Penttilä<sup>2</sup>

Version: 0.2.0 (September 24, 2021)

## Contents

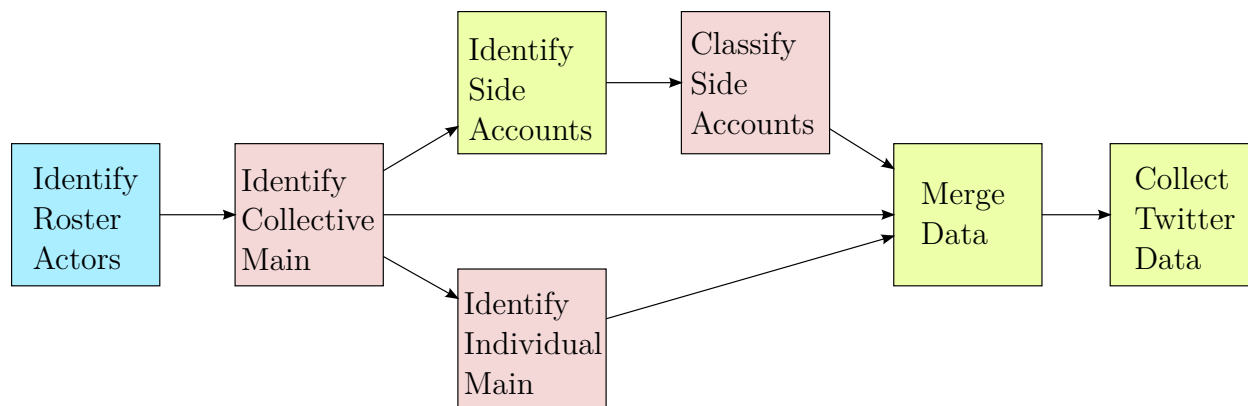
<b>1</b>	<b>Overview</b>	<b>1</b>
<b>2</b>	<b>Protocol</b>	<b>1</b>
2.1	Collective Main . . . . .	2
2.2	Individual Main . . . . .	2
2.3	Collective Side . . . . .	3
2.4	Individual Side . . . . .	3
<b>3</b>	<b>Additional Considerations</b>	<b>4</b>
3.1	Joint Affiliations . . . . .	4
3.2	“Opinions are my own” and Other Disclaimers . . . . .	5
3.3	Temporal Variation . . . . .	5
<b>4</b>	<b>Frequently Asked Questions</b>	<b>6</b>
<b>5</b>	<b>Programming and Other Requirements</b>	<b>8</b>

---

<sup>\*</sup>Contact: Ted Hsuan Yun Chen ([ted.hsuanyun.chen@gmail.com](mailto:ted.hsuanyun.chen@gmail.com)). Authors are listed alphabetically by surname as everyone contributed equally to developing the protocol. This work is funded by Academy of Finland grant nos. 320780 and 320781. We thank Otto Schultz and Amirmohammad Ziaei for providing invaluable research assistance in developing this protocol. Project repository: [github.com/tedhchen/componMultilayer](https://github.com/tedhchen/componMultilayer).

<sup>1</sup>Department of Computer Science, Aalto University

<sup>2</sup>Faculty of Social Sciences, University of Helsinki



**Figure 1:** Data collection pipeline. Blue indicates pre-requirement; red indicates human task; yellow indicates code-assisted task.

## 1 Overview

In this codebook we document the procedures used to collect Twitter account information from organizations in our policy actor roster, including the account information of their members, employees, and other affiliates. We outline a set of standard protocol that aims to guarantee comparability of data collected in different contexts (e.g. countries or policy systems).

## 2 Protocol

We collect the Twitter account usernames of all organizations in our policy actor roster. We also collect the account usernames of the individuals who represent the organizations. This approach affords a more complete picture of an organization’s activities on the Twitter platform – individuals from the organization tend to behave differently from formal organizational accounts but still are recognizable to the public as members of the organization (e.g. Greta Thunberg versus Fridays for Future). The relative comprehensiveness of the data allows us to ask theoretically-motivated questions that were otherwise difficult with only survey data, including the role of formal and informal ties in policy systems and the relationship between individuals and the collective within an organization when it comes to organizational goals.

We specifically identify four “levels” of accounts for each policy actor: 1) collective main, 2) individual main, 3) collective side, and 4) individual side. These levels are summarized in Table 1, using the Financial Times as an example of the roster entry. Detailed instructions for identifying these levels are in the sections below.

Level	Description	Examples
Collective main	Main account of the organization at the national level.	@FT
Individual main	Executive personnel. This level includes only the top-level executive people (at the national level) and the chair of the board of directors. The rest of the board of directors are excluded from this category completely.*	@John_Ridding; @khalafroula
Collective side	Auxiliary accounts of the organization at the national level that represent subunits or focused functions of the organization.	@FTPressOffice
Individual side	Auxiliary personnel that are not members of the executive branch but work for the organization.*	@Mikepeeljourn

\*Additional instructions for specific organization types (e.g. universities) are discussed in the main text.

**Table 1:** Summary of Different Organizational Levels

## 2.1 Collective Main

There is only one account in this category for each organization. The account can be identified in a number of ways, including from the organization’s website, from a Twitter search, or from a general websearch.

In some instances, when an organization is an international one but originated in a specific country, the “national-level” organization will be the international one (e.g. Nokia, which is on the Finnish roster, only has an international account). It is up to the country team to decide whether this is appropriate or to specify that there is no national-level organizational main account.

## 2.2 Individual Main

This level includes the executive personnel of the organization. We consider these as individuals with decision-making power for the entire organization. This specifically means that we want to collect leaders of the top-level unit in the organization (i.e. before units become parallel). Leaders of lower-level parallel units (i.e. subunits) should not be included.

For example, in a business organization individual main accounts would be members of the top level executive management team. If this team includes leaders of subunits, they would classify as individual main accounts. If they are not members of the top level team, they would not be considered as individual main accounts.

Individual main accounts can be identified in a number of ways. If the organizations’ website does not list the executive personnel, LinkedIn serves as an alternative means as the people tend to publish their role in the organization there. Additionally, LinkedIn allows for searching employees of the organizations by relevant keywords (e.g. “executive”).

Type	Rule
Government Ministries	Include the Minister and Minister’s office accounts if available as individual main.
Political Parties	Include only the party chair and vice chairs, the party secretary, and the chair of its parliamentary group as individual main. Include only members of the national legislature or supranational governments (e.g. EU) as individual side.
Universities	Include only rectors, principal deans, and administrative executives (e.g. Director of Communications), or their representatives as individual mains. Do not include students, including doctoral students, either as mains or sides.

**Table 2:** Additional Type-based Requirements for Collecting Individual Accounts

## 2.3 Collective Side

This level includes the subunits/subfunctions of the national-level organization. We have developed a Twitter-specific protocol to collect this data. Specifically, we only include Twitter accounts that satisfy all of the following criteria:

1. Follows the organization’s collective main Twitter account
2. Is followed by the organization’s collective main Twitter account
3. Has at least one of the organization’s pre-specified keywords in its Twitter bio; these organization-specific keywords are specified by the country team but usually include at least different variations of the organization’s name
4. Twitter bio passes human filtering as the organization’s subunit/subfunction

To facilitate this task, we have developed python code that will identify and extract Twitter accounts that satisfy conditions 1–3. The code returns a list of each Twitter account’s username and their bio text. Conditions 4 is checked by a human coder using this preprocessed list (i.e. it is up to the country team to decide which subunits are relevant.) Instructions for using the code is discussed below. The code is available at the project’s Github repository.

## 2.4 Individual Side

This level includes all identifiable individuals that work or are members of the organization but who are not in the individual main level. Because this is a difficult category to collect, we rely on the same criteria as the collection for collective side accounts. Specifically, we only include Twitter accounts that satisfy all of the following criteria:

1. Follows the organization’s collective main Twitter account
2. Is followed by the organization’s collective main Twitter account
3. Has at least one of the organization’s pre-specified keywords in its Twitter bio; these organization-specific keywords are specified by the country team but usually include at least different variations of the organization’s name
4. Twitter bio passes human filtering as the organization’s member or employee
5. Is not already included as an individual main account

Again, we have developed python code that will identify and extract Twitter accounts that satisfy conditions 1–3. Conditions 4 is checked by a human coder using this preprocessed list. Condition 5 is automatically checked by code that merges all data files together. Effectively, this means that we can collect the collective side and individual side accounts at the same time by labelling each entry on the list of username-bio text as collective side, individual side, or neither.

Depending on the type of organization, drawing the line between the ‘main’ and the ‘side’ at the individual level can be difficult. Some organizations have a relatively flat structure without a clear leader or executive team (e.g. social movement organizations like Extinction Rebellion). In such cases, we label all their representatives as executive personnel (i.e. individual main). Some other types of organizations have many voices and executive-like authority occurs at multiple levels. For example, universities can have rectors, deans, professors, and directors of various research centers or institutes. For universities, we identify only administrative executives at the university-level as executive personnel, and we exclude students from both individual main and side levels. Specific requirements for certain organization types are described in Table 2.

## 3 Additional Considerations

### 3.1 Joint Affiliations

An individual may satisfy the criteria for more than one organization. For example, the Chief Executive Officer of Nokia, Pekka Lundmark, is also the Chair of the Board of the Confederation of Finnish Industries. In our data collection task, each organization is collected independently of each other, which means that Lundmark would be first collected as an individual main account when looking at Nokia, then collected as an individual main account

when looking at the Confederation of Finnish Industries. This approach ensures we do not miss any joint affiliations, which can be used in different ways depending on the analysis.

### 3.2 “Opinions are my own” and Other Disclaimers

Some Twitter accounts will include disclaimers such as “opinions are my own and do not represent those of my employers”. We choose to disregard such statements when collecting account data. While some organizations mandate their employees to add a disclaimer, others may also use it strategically to engage in controversial debates without dragging the entire organization along. Whatever the case, if the accounts can be identified using the protocols described here, they are going to be publicly visible members of their organization, and are therefore included in the data. Whether their opinions align with those of their organization is left for the analysis to decide.

### 3.3 Temporal Variation

Twitter is changing constantly. Organizations merge and change their names. Individuals change organizations or end up deleting previous content or even their entire account. Sometimes Twitter suspends accounts, making them unviewable or otherwise inaccessible. Our collection protocol is designed for contemporary data. This means that the closer the times between the policy actor roster construction and the Twitter account collection, the more accurate the data.

However, the temporal dimension is one of the benefits of using Twitter data, as policy actor rosters are often constructed prior to the present. If a country team decides to collect additional data on past affiliations, we describe some approaches here.

To ensure the temporal validity of the Twitter data and the policy actor roster, ideally the individuals are assigned an affiliation that they had at the time of roster construction. Checking who worked for the organization in the past can be done with the [Wayback Machine](#), which shows certain historical versions of websites. However, it takes a while to open archived websites. Figure 2 shows the Wayback Machine interface.

Using LinkedIn is a faster alternative. LinkedIn can be used to identify people who currently work for the organization or to check where an individual worked in the past. Neither Wayback Machine nor LinkedIn contain all relevant information, so country teams can also use local knowledge to fill this information.

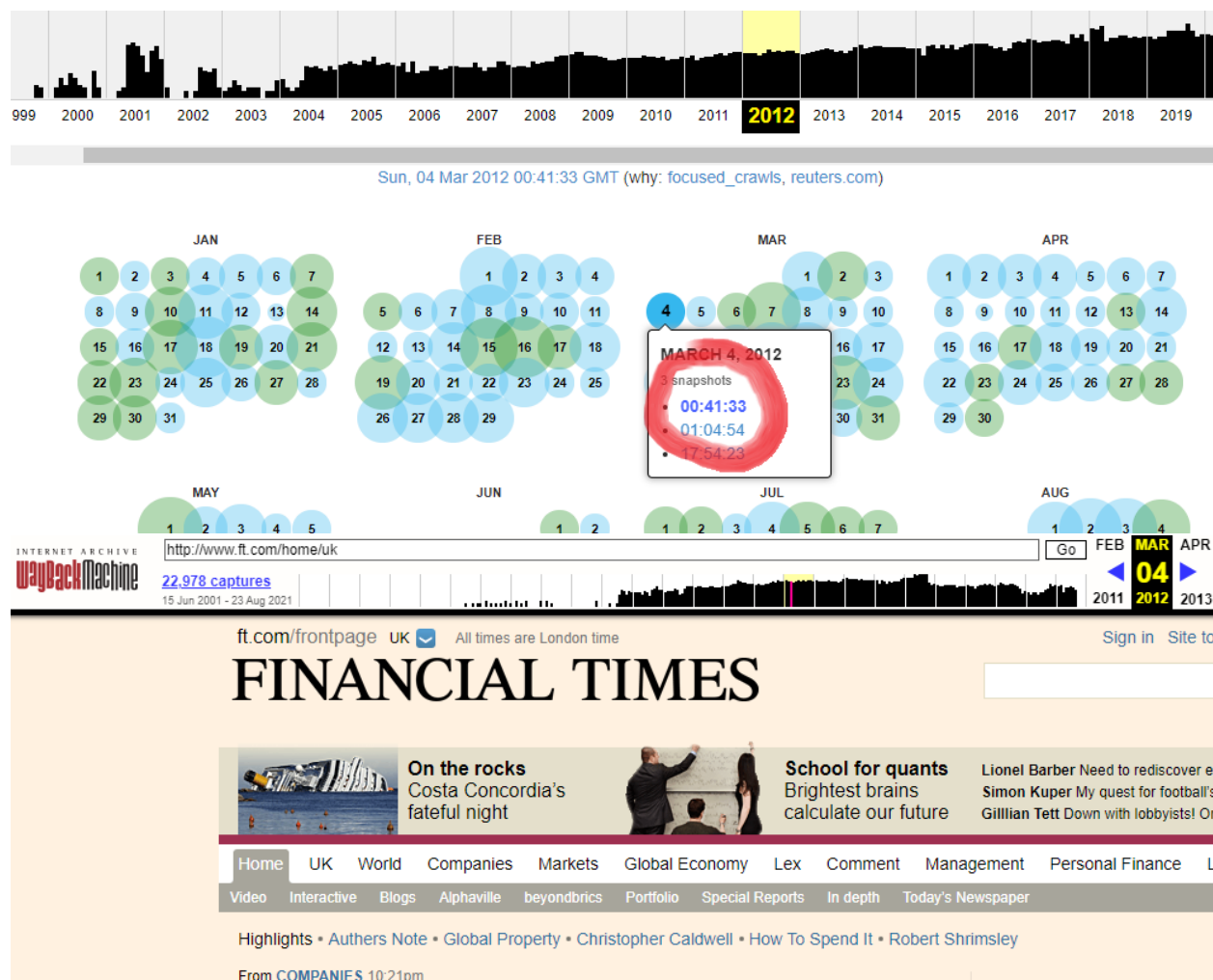


Figure 2: Wayback Machine's Interface

## 4 Frequently Asked Questions

Should an individual who indicates in their bio that they vote for a certain political party or that they share the views of a certain political party be coded as individual sides?

For political parties, stricter rules apply (see Table 2). Specifically, only members of the national legislature (or supranational governments like the European Union) should be coded as individual sides. This means that supporters do not count as individual side accounts. Similarly, even regular members of political parties are not coded as individual sides.

**Should an individual who includes the organization's name in their bio, but their relationship to the organization is unclear, be coded as individual side?**

If there is no certainty that the account is an employee or member of the organization, then they should not be included.

**What if the individual's bio is in a language the coder does not understand?**

When coding the individual side accounts, include only individuals whose bios are in English or a reasonable working language of the country. It is up to the country team to decide what are considered reasonable working languages.

**If an individual indicates in their bio that they previously worked for the organization, should they be included as an individual side account?**

No.

**Do regional party organizations' accounts count as collective sides?**

Regional party organizations (e.g. state or provincial chapters) should not be included because they operate on a different level of public administration than the national party.

**When coding university accounts, are PhD candidates included as individual side accounts?**

Students should not be included as individual side accounts even if they are doctoral students employed by the university.

**In terms of universities and research institutions, there are often sub-accounts for research projects or study programs. Should they be included as collective side accounts? How about accounts for joint projects where multiple organizations are mentioned in the bio?**

The account should be coded as collective side for all the organizations mentioned in the bio. As discussed in subsection 3.1, each organization is collected independently of each other, so if a research project account shows up for multiple organizations, it should be included as an individual side account for every organization they are affiliated with.

**How should national branches of international organizations be coded?**

The national account of the policy actor should be coded as collective main, except in cases where the national and international accounts are the same (e.g. Nokia in Finland). Only subunits/subfunctions should be coded as collective sides, meaning that other national units should never be included.

For example, when coding Greenpeace Finland's accounts, Greenpeace International or Greenpeace Spain should not be counted as collective sides, nor should be their employees or members be counted as individual sides.



## **5 Programming and Other Requirements**

[Coming soon]