It's what's not mentioned that flashes: Twitter analysis of UN agencies' timelines

April, 2022

Main takeaways

- Topics specifically related to the development of statistics and data almost never show up in the Twitter feeds of specialized UN agencies.
- Even though statistical agencies have pushed for more funding in gender data since 2019, this did not lead to increased attention by major UN agencies.
- Partner agencies focused on the development of statistics and data have increased their efforts to highlight relevant topics over recent years.

Over the past years, data financing partners around the globe have tried to highlight the importance of well-funded statistical systems especially in low- and middle-income countries to track progress along the SDGs and inform impactful policy. This campaign is put forward through multiple channels with social media arguably reaching the widest audience. With multiple platforms avaiblable nowadays, the most commonly used among those for purposes of communication in the international community has been Twitter. But how well are statistical agencies and partners doing in generating attention for their pivotal cause on Twitter? And how much attention do non-statistical agencies pay to the relevant question of development in data and statistics?

---

1 Scope

To see how much attention is payed to the development of data and statistics in low- and middle-income countries by central institutions of international cooperation, the Twitter timelines of the 17 specialized UN agencies were analyzed, as for example *FAO, IMF, IMO, UNDP, UNESCO, WHO, World Bank, WMO* and *WTO*. Some of them do exhibit a statistics and data division that was however included into the list of 22 specialized institutions and think tanks that focus primarily on statistics and data. The list comprises for example *PARIS21, Open Data Watch, FAOStat, Royal Statistical Society, Eurostat* or the *Global Partnership for Sustainabele Development Data* (see the full list in the appendix).

---

2 Approach

Twitter offers an easy to use access through an API which allows to fetch up to 3200 of the most recent tweets from timelines of users. Depending on the tweet frequency of the agency or institution analysed, this reaches back up to 2018. Overall, we gathered a total of almost 67000 tweets for UN agencies and a little more than 48000 tweets for statistical agencies. For the actual analysis however, we resticted ourselfs to the

hashtags that were used in these tweets, instead of the text of the tweet itself. The rational stemms from the observation that usually hashtags are used to condense the most important information into one marker. Therefore, these hashtags provide a rich set of information about topics that UN agencies want to highlight and will serve as a basis for this analysis.

The threefold analysis focuses at first on differences of the twitter activities of UN agencies and statistical institutes and reveals most importantly what can't be found in the timelines of UN agencies. Just as interesting as completely absent trending topics on the part of UN agencies, are frequently mentioned issues spotlighted by statistical agencies. In the second part of the analysis, the UN agencies will be put under scrutiny regarding their mentions of statistical topics. These will be defined by a list of key hashtags that are chosen according to their overrepresentation in statistical agencies' Twitter feeds. Lastly, we will have a look at the time evolution of the most frequently used hashtags by statistical agencies.

---

3 Analysis

3.1 Differences in trending topics for UN agencies and statistical agencies

Before diving deeper into mentions of statistical topics in twitter feeds, it is helpful to get an overview of the relation of the frequently mentionted sujects. The Venn digram in Fig. 1 shows that statistical agencies and UN specialized agencies share only 14% of the overall hashtags used which does already hints at the mild usage of statistical hashtags by UN organizations, if one assumes that statistical agencies tweet predominantly about statistics-related topics. Note, that the absolute numbers in the Venn diagram give the number of distinct hashtags and explicitly not the summ of the occurences of these distinct hashtags.

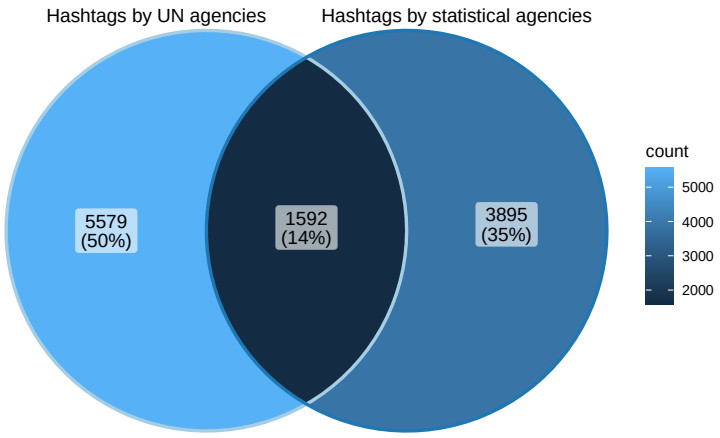


Figure 1: Venn diagram of the all hashtags that were used by UN and statistical agencies. Only 14% of hashtags are used by both groups of agenices.

Having a closer look at the twitter activity of statistical agencies to get a better impression what common hashtags they share, Fig. 2 shows the commonalities of three important partners in statistical development. Interestingly, the distinct hashtags they share make up only 9% of the all utilized hashtags. Zooming in on the intersection, the Word Cloud in Fig. 3 shows the hashtags and therefore topics that show up in twitter feeds of all three organizations.

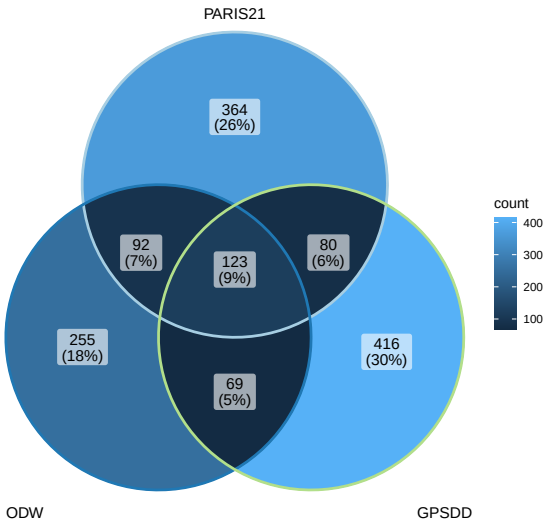


Figure 2: Venn diagram of hashtags that are shared by the three important partners *P21, ODW* and *GPSDD*.

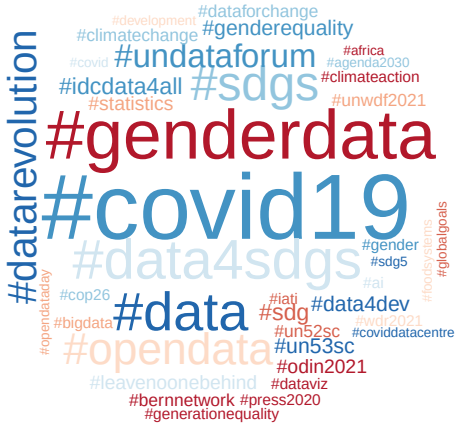


Figure 3: Word cloud of the intersection of *P21, ODW* and *GPSDD*. Note that the size of the words scales with their frequency in the set.

hash	Freq_Stats	Freq_UN
#dataforgood	634	0
#undataforum	554	0
#datascience	530	0
#datarevolution	368	0
#sdg4data	259	0
#odisummit2021	253	0
#datavalues	223	0
#crvs	217	0
#odisummit2020	210	0
#statstrivia	192	0

hash	Freq_Stats	Freq_UN
#genderdata	1756	1
#data4sdgs	600	2
#faostat	214	1
#sdgreport	174	2
#wdr2021	159	2
#betterdata	158	1
#r	112	1
#eudatathon	74	1
#pinksheet	74	2
#dataforchange	65	1

Table 1: The 10 most frequent hashtags that show up only in Twitter feeds of statistical agencies but never in UN agencies on the left. On the right, the 10 most frequent hashtags that only appear once or twice in UN feeds.

Now to the most striking fact: Even though agencies specialized on statistics and data have been calling for intensified efforts to strenghten development of statistics around the globe, this has not reached the attention of the major UN agencies. Hashtags like `#datarevolution`, `#dataforgood` or `#undataforum` never showed up, `#genderdata` or `#betterdata` only once. All this despite an ever growing focus on these topics in recent years (compare Sec. 3.3).

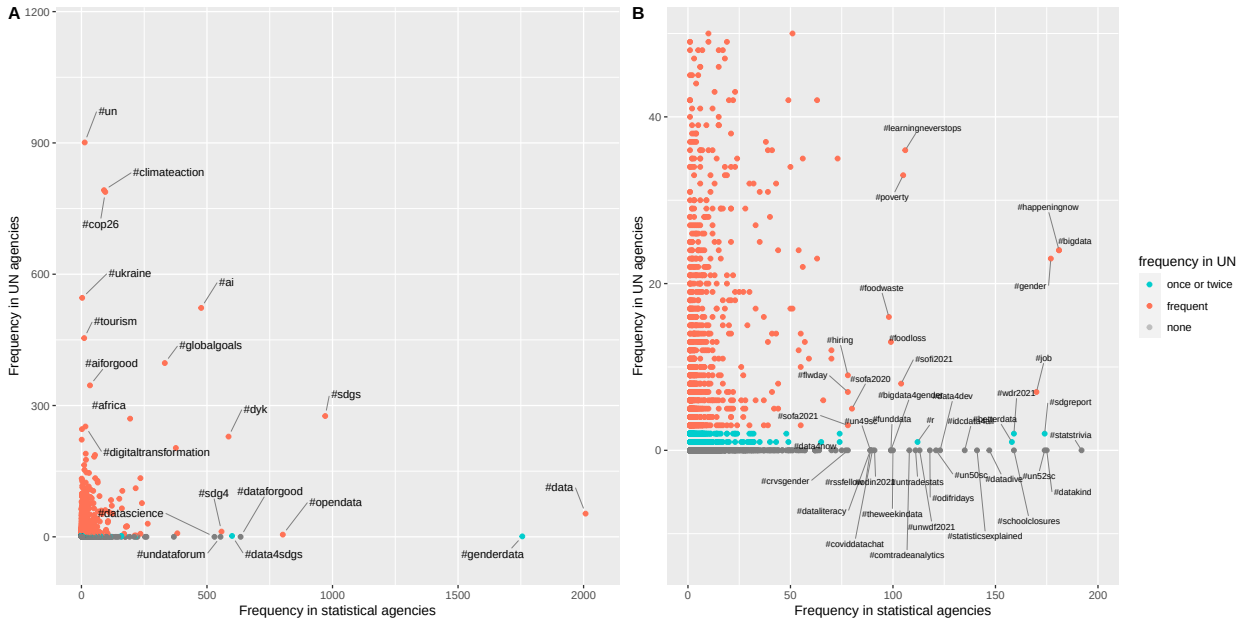


Figure 4: The hashtags used in Twitter feeds of the UN and statistical agencies along the two dimensions of frequency in either category. Graph A on the left shows all hashtags except the outlier *#covid19* which had over 3200 occurences in both categories. In graph B on the left, it was zoomed in to illustrate the low usage of statistical hashtags in UN twitter feeds.

This lack of statistical topics on the frontpage of major UN agencies becomes even more apparent when the frequencies of hashtags in both groups is juxtaposed. On the left in Fig. 4A, hashtags are displayed along the dimensions of frequencies in both groups of agencies and the differing foci are clearly visible: While statistical agencies focus understandably more on topics of statistical development, UN feeds diplay rather other current topics. This theme is just as pronounced when zooming in on less frequnently used hashtags in Fig. 4B where statistics hashtags exhibit a much lower frequency on the UN side of the Twitter sphere.

3.2 References to statistical development in UN agencies' twitter feeds

Turning now to the use of statistics by UN agencies, the key statistics hashtags are identified according to the following steps:

(i) First, the frequencies of each hashtag used by UN agencies is computed (intersection in Fig. 1).
(ii) Secondly, a list of hashtags related to statistics is determined by comparing the `odds ratios` of hashtags.

- The odd ratio of one hashtag is the relative frequency of this hashtag among all tags used by Stats Think Tanks on the relative frequency of this hashtag in the twitter timelines of UN agency.
- Mathematically, this means for #x:

$$\frac{Frequency\ \#x\ in\ Stats\ Think\ Tanks\ tweets/Total\ number\ of\ \#s\ \ in\ Stats\ Think\ Tanks\ tweets}{Frequency\ \#x\ in\ UN\ agencies\ tweets/Total\ number\ of\ \#s\ \ in\ UN\ Agencies\ tweets}$$

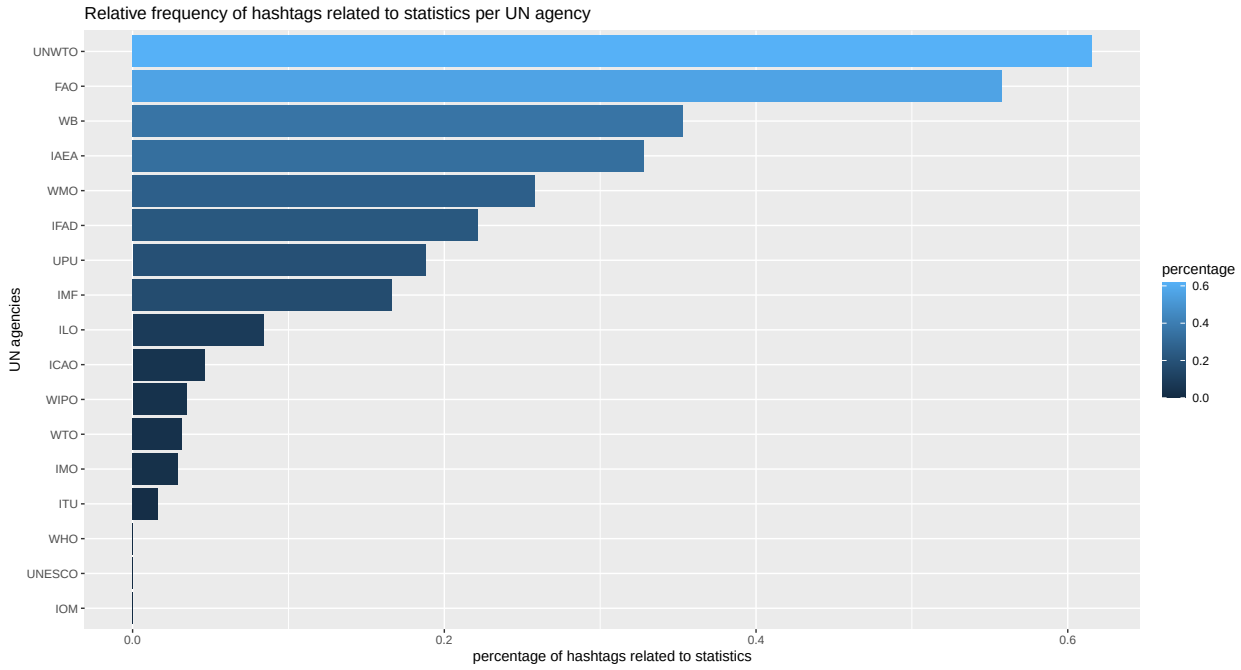
(iii) All hashtags that exhibit a frequency lower than 5 are excluded since they wouldn't countribute to the analysis if they rarely appear in UN twitter feeds.
(iv) The 20 hashtags with the highest odds ratios were chosen as the final key hashtag list.

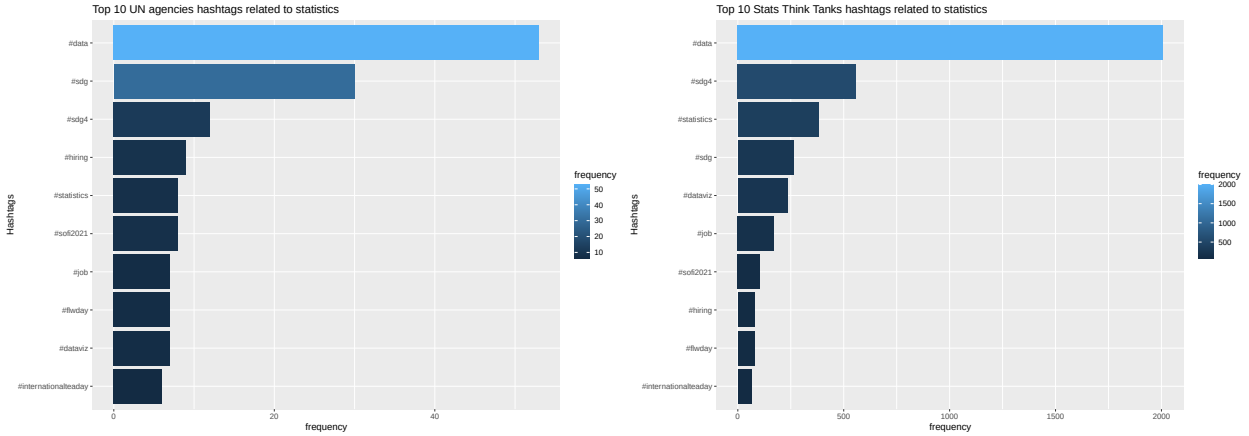
The final list is found below. The outlined procedure seems justified for a definition of key statistical terms, since a high odds ratio meas that they show up way more often on the Twitter feed of statistical agencies than on UN agencies.

hash	Freq_Stats	Freq_UN	odds
#statistics	382	8	66.255153
#sdg4	558	12	64.520725
#data	2008	53	52.569533
#dataviz	235	7	46.581783
#job	170	7	33.697460
#sofi2021	104	8	18.038052
#flwday	78	7	15.461187
#internationalteaday	66	6	15.262967
#sdg	264	30	12.210374
#hiring	78	9	12.025368
#gender	177	23	10.678044
#foodloss	99	13	10.566669
#bigdata	181	24	10.464383
#happeningnow	181	24	10.464383
#csw66	70	11	8.829816
#digitalagriculture	37	6	8.556512
#foodwaste	98	16	8.498698
#teachers	41	7	8.127034
#healthydiets	70	12	8.093998
#earthobservation	33	6	7.631484

The above word cloud shows the most important hashtags in feeds of statistical agencies that are on the key hashtag list.

To show the refferences to statistical development in UN Twitter feeds, the relative frequency of hashtags related to statistics is computed for each UN agency, based on the hashtags identified through the process using odds ratios. The relative frequency of hashtags related to statistical development for each agency is simply equal to the frequency of hashtags related to statistics used by each UN agencies over its total number of hashtags. The result for each agency is reported below. Again, it is evident that the usage of hashtags related to statistics is fairly low among all UN specialised agencies.





3.3 Trending of the top hashtags overtime

The top 6 hashtags used by the statistical agencies are `#covid19`, `#data`, `#genderdata`, `#sdgs`, `#opendata` and `#dataforgood`. By creating the graphs of their trending history, we found that different hashtags evolve differently overtime. The frequencies of some hashtags show clearly sharp increases during global summits and decreases afterwards. Frequencies of other hashtags, on the other hand, increase steadily over time. The hashtag `#covid19` was the most used hashtag by Statistical Agencies despite only being used since March 2020. It shows the statistical community was respoding rapidly to the emerging crisis.

4 Limitations

This methodology has several limitations.

a. Tweets are a non-representative data source made available by a private company. The methodology is therefore subject to biases and depends on continued access to the Twitter API, which may not be sustainable in the long term.
b. Hashtags usually do reflect central topics of a tweet but tweets can nontheless avoid hashtags wich would not allow to include them with this methodology.
c. Difference between frequency and impact of topics: Some topics may trend on Twitter and even induce herding behaviour among the agencies but do not make it on the agenda of high-level events.

d. Using hashtags with the highest odds ratio to define the key statistical hashtags list might non render the hashtags most relevant to development in statistics and data since a high odds ratio merely reflects an overrepresentation in statistical agencies' twitter feeds.

Appendix

All Specialized UN agencies used in this analysis are displayed in the upper table, whereas all statistical agencies and think tanks are shown in the lower table.

twitter	acronym	name
FAOnews	FAO	Food and Agriculture Organization of the UN
iaeaorg	IAEA	International Atomic Energy Agency
icao	ICAO	International Civil Aviation Organization
ifadnews	IFAD	International Fund for Agricultural Development
ilo	ILO	International Labour Organization
imfnews	IMF	International Monetary Fund
imohq	IMO	International Maritime Organization
itu	ITU	International Telecommunication Union
unesco	UNESCO	UN Educational Scientific and Cultural Organization
upu_un	UPU	Universal Postal Union
who	WHO	World Health Organization
wipo	WIPO	World Intellectual Property Organization
worldbank	WB	World Bank Group
wmo	WMO	World Meteorological Organization
unwto	UNWTO	World Tourism Organization
wto	WTO	World Trade Organization
UNmigration	IOM	International Organization of Migration

twitter	acronym	name
ContactPARIS21	PARIS21	PARIS21
OpenDataWatch	ODW	Open Data Watch
worldbankdata	World Bank Data	World Bank Data
FAOstatistics	FAOSTAT	FAO statistics
Data2X	2X	Data2X
DataKind	datakind	DataKind
DataLook	datalook	DataLook
datapopalliance	Data Pop	Data-Pop Alliance
od4_d	od	OpendData for development
UNESCOstat	UIS	Unesco Institute for Statistics
UNICEFdata	UNICEF Data	UNICEF Data
TuvaLabs	Tuva	Tuva Labs
Globaldatalab	GDL	Global Data Lab
DataInnovator	DI	Data Innovator
RoyalStatSoc	RSS	Royal Stat Society
DFID_Stats	DFID Stats	DFIF Statistics
IntStat	ISI	The International Statistical Institute
UNTradeStats	Unstats	UN Trade Statistics
ODIHQ	ODI	Open Data Institute
EU_Eurostat	Eurostat	Eurostat
UNStats	UNSD	UNSD
Data4SDGs	GPSDD	Global Partnership for Sustainable Development Data