Mapping OSINT Landscapes →

→ Research Question

How can we explore OSINT landscapes through Twitter?

→ Summary

In this exploratory project, the guiding questions have descriptive goals that focus on different aspects of OSINT practice. We build upon a dataset of Tweets including the #OSINT and experimented with multiple ways of visualizing what the actors assembling in it are doing. We reflected upon networks of actors, images, methods and issues.

4. Issue mapping express

→ Summary & Findings

Our research focused on mapping the issues discussed surrounding the Open Source Intelligence (OSINT) space on Twitter. OSINT is becoming an important tool in building and influencing public narratives. Understanding how the space evolves over time will help track use and misuse of the tool as the discussion continues to flourish all over social media.

OSINT as a topic on Twitter has grown tremendously since the war against Ukraine begun in February of 2022. As seen in the histogram below there is a clear rise in the number of tweets around this time. Additionally, conversations being had in the space differ before and after the war began. For example, before the invasion announcement, "russia" was discussed in reference to OSINT tools used to decipher general misinformation, such as identifying Russian-led websites. However, after the invasion announcement, the OSINT community focused on tracing and verifying Russia's claims in the context of the war against Ukraine.

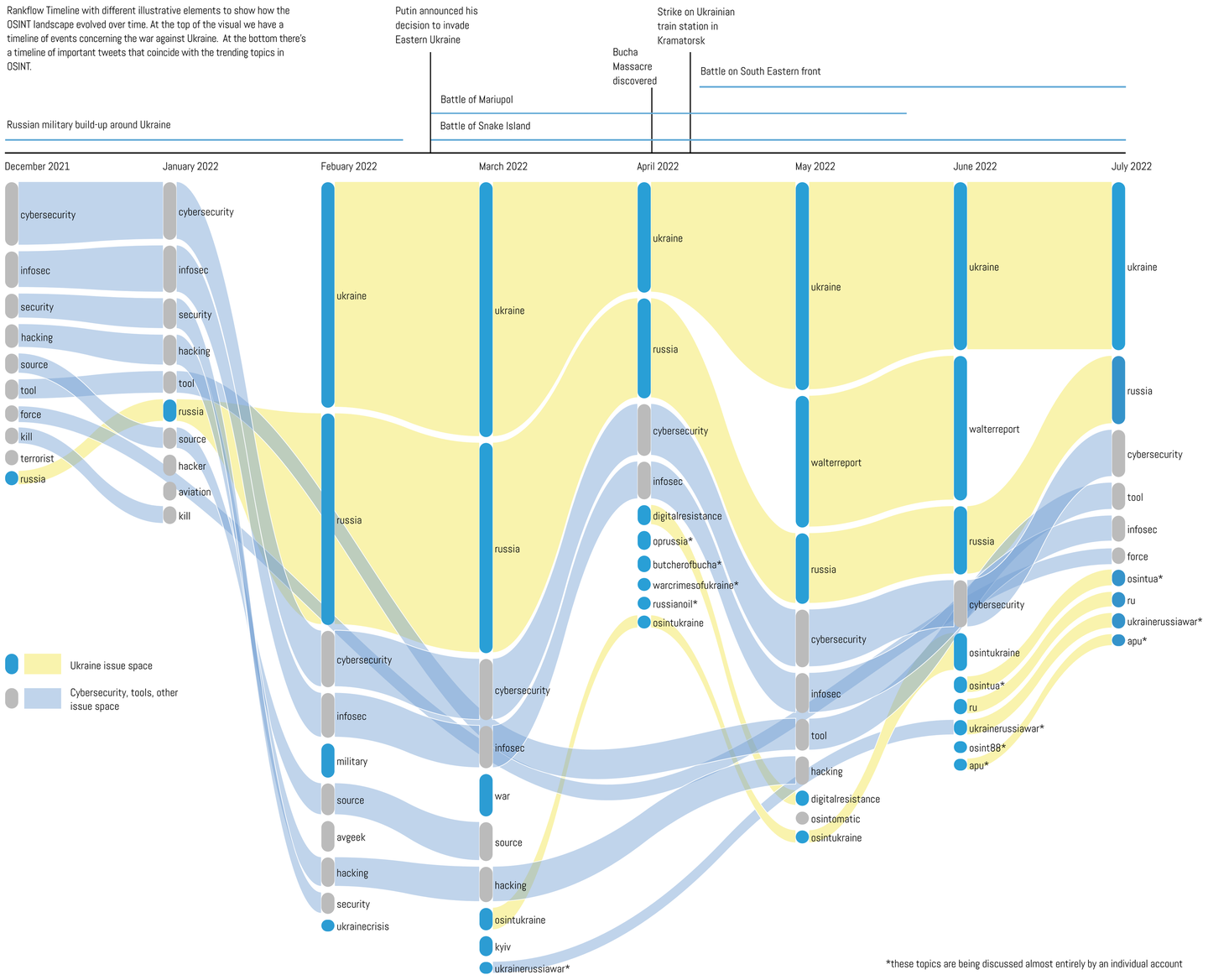
The Rankflow Timeline we created below shows this even more clearly, as we see topics surrounding "cybersecurity" and "infosec" leading the conversation, but then in February 2022 "ukraine" and "russia" are far and above the most talked about topics. There is a huge spike in the number of tweets after Putin's announcement regarding the invasion of Ukraine. We see the most active/popular actors change from blogs/organizations to individuals and the discussions are about using OSINT to share information regarding the war against Ukraine rather than general discussion about the topic.

While the number of tweets under the #OSINT hashtag has clearly increased after the invasion, so have the number of individuals spamming tweets. These passionate users tend to muddy the data collected with specific hashtags and hashtag groups that aren't necessarily representative of the entire #OSINT landscape. For instance one of the most important terms we found after running a term frequency–inverse document frequency (tf-idf) function on our data was "russianoil" which was a specific topic discussed by an individual user, RuOilTracker. In the Rankflow Timeline we note other examples of important terms that show up in our data due to passionate individuals in the space.

→ Research Questions

1. How did the #OSINT Twitter landscape evolve in light of the war against Ukraine?
2. What are the most prevalent issues connected to #OSINT?
3. Did the discussion around #OSINT grow after the invasion of Ukraine?

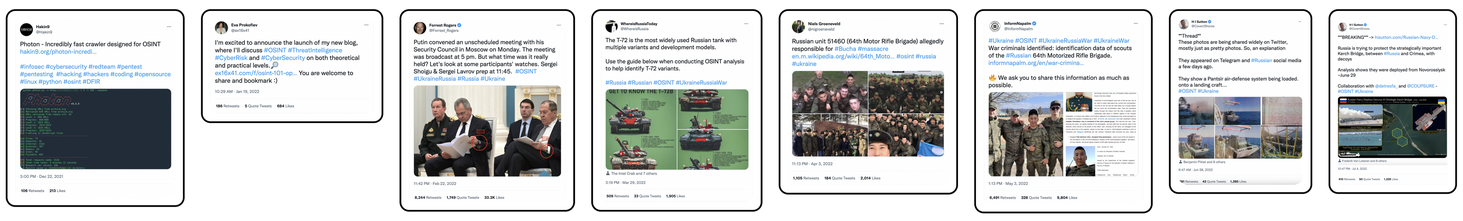
Rankflow Timeline with different illustrative elements to show how the OSINT landscape evolved over time. At the top of the visual we have a timeline of events concerning the war against Ukraine. At the bottom there's a timeline of important tweets that coincide with the trending topics in OSINT.



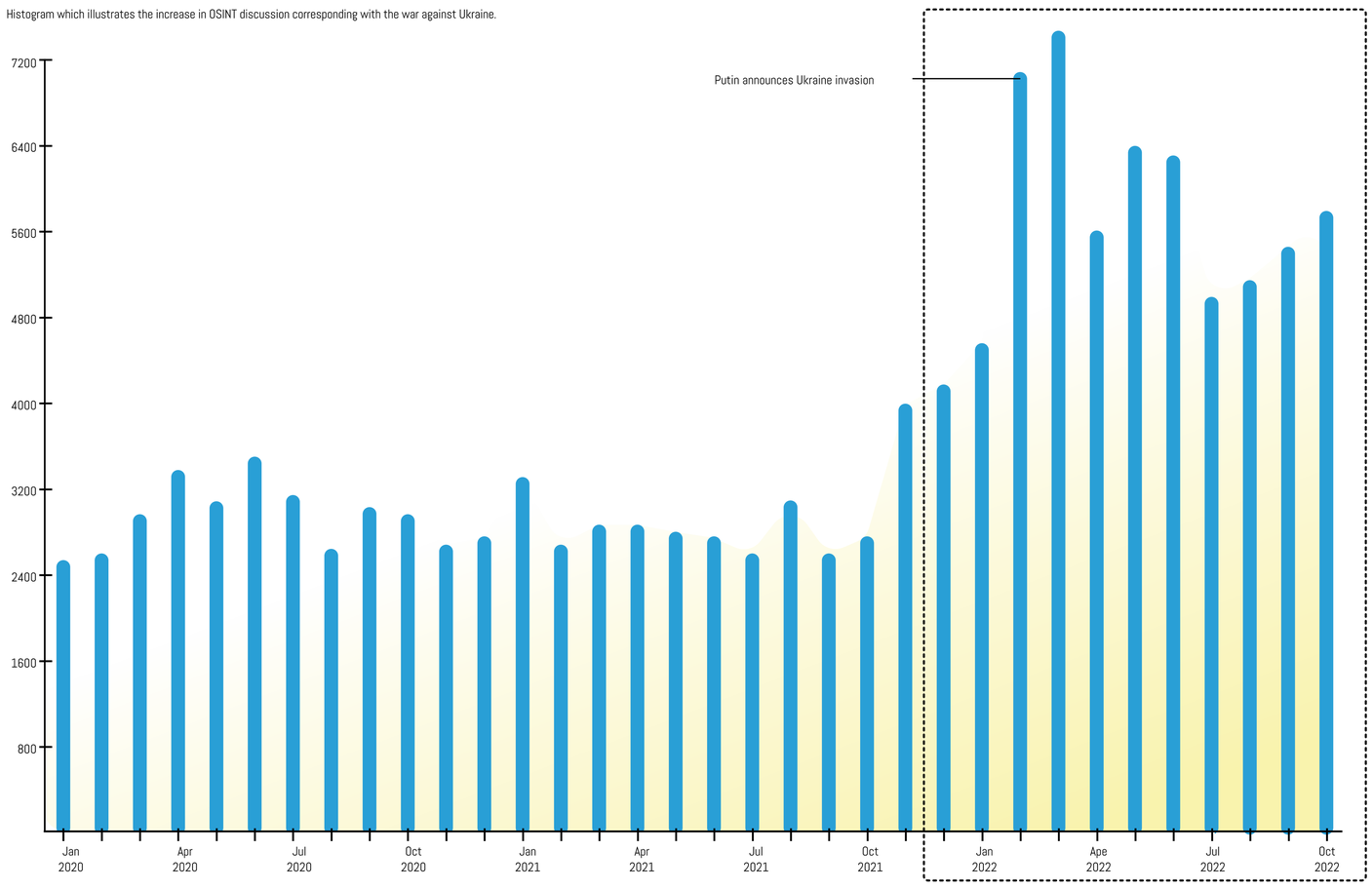
*these topics are being discussed almost entirely by an individual account

Word Cloud which shows the biggest topics discussed from December 2021 to July 2022.

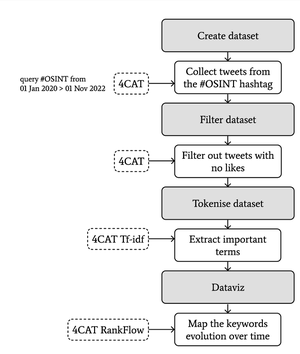
Meme illustrating how dynamic the #OSINT landscape is



Histogram which illustrates the increase in OSINT discussion corresponding with the war against Ukraine.



→ Protocol



DATA COLLECTION
Our initial dataset contains tweets scraped from Twitter using 4CAT. These tweets contain #OSINT in their body and were tweeted between January 1, 2020 and November 1, 2022. This initial dataset contains 990,055 tweets, which we filtered further with a focus on engagement.

We created a rankflow diagram on this filtered data and found that the OSINT issue space changes drastically right around the time of Putin's announcement regarding invading Ukraine. We created a timeline of events during the few months following the start of the war. This timeline helped us connect and contextualize the increase in engagement and change in issues discussed online which match events on the ground.

After applying the context of the war we investigated this dataset further and looked at the most popular tweets and users to better map the OSINT landscape. Some of these tweets can be seen in the Rankflow Timeline on the left.

DATA CURATION
An issue is a topic for debate or discussion, but our initial dataset contained unfiltered OSINT tweets regardless of engagement. We investigated the dataset and found that there were bot accounts that would tweet thousands of tweets with no engagement (e.g. the account @DataAbyssAI) adding no meaningful discussion to the topic. We filtered the initial dataset further using likes as a metric, filtering out any tweets that received no likes, as engagement is an essential part of an issue space (Rogers et al., 2015).

Using 4CAT to manipulate our filtered dataset, we tokenised the body of tweets to run Natural Language Processing queries and visualize our findings. We ran multiple iterations of tokenisation, pulling out certain terms that we deemed too broad or associated with spam (e.g. emojis and non-latin alphabets). After curating our data we noticed that the #OSINT issue space was mainly focused on discussions about cybersecurity and tools up until the invasion of Ukraine when the war became the top issue being discussed within the space. We decided to focus our investigations into the data on the time period between December 2021 and July 2022 where the most significant changes in the landscape occurred.

VISUALIZATION AND ANALYSIS
To visualize our findings, we generated 3 different graphics using 4CAT and other software. Our first visualization used 4CAT to generate a histogram showing the number of tweets per month in our dataset of liked tweets regarding #OSINT. This visualization shows how the discussion around OSINT increased dramatically after the invasion of Ukraine.

Our second visualization used a combination of 4CAT and RankFlow, a tool by Bernhard Rieder that we found online to create a rankflow diagram. First we found the top 10 most important terms for each month during our chosen time period using 4CAT. Then we counted how many times each term appeared in the separate tokenised text files. Finally we took this data and put it into the RankFlow tool to create our rankflow diagram. This visualization shows how the discussion around OSINT evolved as events surrounding the war against Ukraine occurred. We can see the exact month where the majority of the conversation changes from cybersecurity and OSINT tools to the war against Ukraine. To add more detail we created a timeline of events and placed it above the rankflow diagram to show what was happening outside the Twitter space to influence discussions surrounding OSINT. We then picked out the most popular/interesting tweets and created another timeline below the rankflow diagram.

Lastly, our final visualization used a combination of 4CAT and a free Word Cloud Generator we found online. We combined the tokenised text files of our chosen time period, counted the number of times each one of the top monthly terms appeared, then entered them into the online generator. This visualization gives a quick overview on the issues discussed during our chosen time period between December 2021 and July 2022.

FACILITATORS:
Guillén Torres; Marc Tuters; Lonneke van der Velden

PARTICIPANTS:
Jasmin Shahbazi, Agathe Bourdarias, Kiana Saidah, Marsha Batubara, Tomi Fischer

Global Digital Cultures
University of Amsterdam