7. Monitoring

TL;DR Monitoring

- Monitor groups, narratives and artifacts (e.g. articles on URLs) in the team's area of interest, so when an incident starts, there's a body of knowledge supporting it.
- Identify available relevant datasets in existing collections and social media feeds
- Use searches and APIs
- Make data available to your team in a format they can use
- Watch for biases, and check that your data is clean

Monitoring

When an alert comes in, the incident workflow starts (see the next next chapter for details of this). But whilst the sexy incident responses and "took down a huge operation" responses exist, incident response isn't all that a disinformation response team does.

We build out our knowledge bases and communities, write code to speed up our responses, test tools and processes, and work on a lot of the background things that help an incident response run smoother.

Workflows support the top-level mission and goals of the group. Activities to support those goals include incident response, but they also include:

- monitoring for public safety issues that we can report before they become harmful;
- mapping harmful narratives as they emerge;

- monitoring known disinformation feeder channels.
- chasing down disinformation tactics and counters;
- Adding and maintaining supporting disinformation data

Monitoring work includes spreader analysis - looking for infrastructure and accounts that are set up in advance of incidents, including sock puppet accounts “laundered” and left to mature.

Monitoring Alert Feeds

A team has many places it can potentially get disinformation alerts from. These include:

- Alerts from disinformation team members
- Feeds from other groups
- Phone honeypots
- Reporting hotline (dedicated email address)
- Sniff disinformation report lists, dashboards and botnet feeds for themes
- Set up reporting from social media (Facebook, twitter etc)
- Ask social media companies for feeds from them
- New data coming into the DKAN

We learn about potential incidents from several places:

- Teams connected to this one, e.g. Covid19activation and covid19disinformation, who are watching for disinformation online
- Team members spotting online disinformation and raising the alert in the team slack channel
- Team members spotting alerts from other disinformation tracking teams
- Other CTI channels telling us about disinformation in their feeds

Important: An alert isn't the same as an incident. It's an indicator that something might be worth investigating and starting an incident response for.

Monitoring Narratives

Narratives are part of incidents - each incident might have multiple narratives involved, or just one, but there's usually an identifiable narrative somewhere in there, that you can use to see if there are related incidents already tracked or dealt with etc.

The other thing about narratives is that they, like incidents, have lifetimes. Some narratives appear as a result of a world or local event (or upcoming or anticipated event), and are only useful whilst that event is in peoples' minds. Example: using the Stafford Act to make everyone stay indoors was a narrative we tracked a month ago, before the stay-at-home orders started and it was a lot clearer about what states could, couldn't, would and wouldn't do.

Other narratives appear for a while, go dormant, then reemerge in different forms. Example: 5G, which was originally part of the "radiation of all forms will do bad things to you" narratives, and has now come back in a mixup with covid19.

Organising Narratives

What we need is a way to log all the narratives that we know (or care) about, whilst keeping a smaller list handy of "currently alive" narratives that we can check incoming disinformation against.

There are a lot of narratives: we've seen hundreds of them in Covid19 alone. We've also seen that these can be grouped into themes; we've used mindmaps to group and organise

narratives into hierarchies, making them easier to read and manage. We've also used spreadsheets to share narrative lists.

Identifying new narratives

Part of our work is to identify new threats before they become widespread. One way to do this is to identify emerging narratives from our existing asset collection.

First, we need to establish a baseline understanding of the current threat landscape in our area of interest (e.g. anti-mask, covid5g etc). The places we look to start this work include:

- Master narratives lists
- Existing lists of persistent threats known to carry disinformation: known bots, sources (e.g. disinformation websites), and canaries (accounts or hashtags with a high probability of carrying disinformation in this area)
- Regular threat streams: known disinformation feeds, subscriptions and platforms.

Once we have a baseline, we can establish persistent and repeatable monitoring:

- Identify data sources to monitor, e.g. googlenews, twitter, facebook, news aggregation sites etc
- Create saved or formatted searches for each platform, e.g. twitter = '#disinformation covid qanon boogaloo'; google = google hack formatted with a time parameter, e.g. 'disinformation and covid when=1d'
- Where api access is difficult, use other platform collection resources where possible, e.g. tweetdeck, crowdtangle

Other ways to find outlier or new narratives include watching for one or more of:

- merging and/or reemerging narratives being pushed by usually opposing groups, or old narratives that are reactivating
- local or world events, e.g. protests, changes in an area's status around specific dates (holidays etc)
- anomalous or significantly-sized online activity, e.g. in trending hashtags

Once narratives are found, you'll need to analyse them:

- evaluate source biases (is this state-owned media, an opinion article, social media etc)
- find additional sources with the same and/or competing narratives
- compare and contrast your findings: what's the same - is this fact or opinions? What's different - why? What's the intent and/or agenda behind the narrative - is it political, influence, harm, designed to confuse, distract, disrupt?
- How could this be used for bad (you might want to red team this)
- What would the impact be if this narrative is leveraged for bad?

Disinformation Data

Beware of bias when you use datasets collected by other people. Their collection isn't your collection: be aware of biases, data gaps etc.

Where does Disinformation Data Come From?

The cynical amongst us would say that we're drowning in disinformation data. Mainstream news has many stories about disinformation incidents and takedowns, political groups are quick to decry "fake news", and almost everyone working on disinformation has a favourite fake cure or conspiracy theory.

In practice, if we're looking for disinformation in our specific areas of interest \(e.g. the CTI League currently works on Covid19 related disinformation\) in time to make a difference to its effects, we need to do some groundwork and build out connections, information feeds and catalogues of good places to look.

Types of data

We need to think about data. Mostly we're dealing with data that's moving, at rest and static.

- Moving data: A lot of research places have social media listening - downloading all the social media messages etc around topics, hashtags etc of interest.
- Data at rest: this is the data we've grabbed during investigations, usually as part of finding more of a network and its effects. We're often actively analysing it, working out how we can affect the environment it's in.
- Static data: this data isn't going to change. Some of it is moving data that we've stored, and the environment it was in has been overtaken by events. It's of interest because it contains patterns to be mined, and could contain clues to later behaviours. Other static data is used to support investigations.

Data inputs: Alerts and Canaries

We receive alerts about possible disinformation incidents from members of the disinformation team, and from other teams connected to us. Typically we get alerts around an artefact or theme, e.g.

- A new narrative emerging online, either in general social media or known conspiracy / extremist / target etc groups
- A local or world event that might spark a disinformation incident

- Anomalous or significant-sized online activity that might be associated with a disinformation incident
- Command signals from known disinformation groups (e.g. qanon)

The types of artefact that we typically receive include:

- Images
- Messages, e.g. tweets, facebook posts, SMS or Messenger/Telegram etc messages
- URLs

The processes for investigating these are discussed in more depth in the next chapter.

Several accounts and groups are either known producers or early adopters of many disinformation campaigns. We've dubbed these "canaries", as in the entities that give the first signals that something is happening: canary, as in "canary in a coal mine".

Data sources: disinformation data streams

When we get our first data inputs, it's a good idea to check them against other disinformation and related data collections, to see if they've been picked up by other researchers, or those researchers have already collected data related to these inputs that can be of use to our investigation. The data feeds are continually updated, so are a good source for breaking data; the static data collections are good for finding history on data, source, narratives etc.

Collecting your own data using tools

The datastreams above will help you get a sense of what's known about the artefact and/or theme that you're investigating, and sometimes that's enough to craft a response, e.g. if there's a WHO page on a known scam, that might be enough evidence to ask for takedowns

etc. But most of the time, you'll have to go collect your own data from across social media, and sometimes beyond, e.g. for paper flyers, we asked people if they'd seen them in their neighbourhoods too.

Where you collect from, and what you collect will depend on the artefacts you found, but here are some of the ways.

Twitter data

Twitter data is studied a *lot\precisely because it has a lovely API. Since we use a lot of Python here, let's talk about Python libraries. If you have twitter API codes, then Tweepy is a good choice. If you don't want to use the twitter API, try Twint.

Various researchers post twitter data-gathering tools online. Andy Patel's twitter-gather is good if you're doing twitter network analysis. We have code based on an early version of this in the github repo. It's andy_patel.py - call it with "python andy_patel.py name1 name2 name3 etc" where name1 etc are the hashtags, usernames, phrases \(phrases in quotes\) that you want to search Twitter for. Andypatel.py creates a set of files in directory data/twitter/yyyymmddhhmmss_hashtag1 etc with the tweets, most prolific urls, authors, influencers, mentions etc and gephi input data so you can create user-user etc graphs (see the gephi instructions in this BigBook for how to do that\). Data for earlier investigations are in the repo folder data/twitter if you want to see what that looks like.

Useful references on collecting twitter data include https://firstdraftnews.org/latest/how-to-investigate-health-misinformation-and-anything-else-using-twitters-api/

Facebook data

The Facebook API is horrible. Most everyone tracking social media uses a third party like CrowdTangle or scrapes for the data they want. The Crowdtangle chrome extension is available free to anyone, but the full Crowdtangle tool isn't: it's available to news organisations, some academics, and pilot programs, so it's worth checking to see if your team is eligible or has someone with access on it.

Reddit

Reddit data is regularly dumped in an easy to read format. For quick-looks, there are tools like (https://www.redective.com/)

Multi-platform tools

Reaper collects from a set of social media feeds. Trying that out. If you have issues with Facebook access tokens, look at list in https://developers.facebook.com/docs/facebook-login/access-tokens/ - then used https://developers.facebook.com/tools/explorer/ to check the token worked before putting into Reaper. If you get “Page Public Metadata Access requires either app secret proof or an app token”, see https://developers.facebook.com/docs/apps/review/feature#reference-PAGES_ACCESS

Storing datasets

A tracking team will collect a lot of supporting data that isn't artefacts: things like the tweets and accounts associated with a hashtag, or urls and groups that a story appears on. Most of this data isn't part of reports - it's supporting data - but still needs to be stored somewhere, for analysis.

Social media data can be large, and its value is often in the relationships between objects as well as the objects themselves. Options we've used include collections of CSV and json files held in a DKAN data warehouse, Neo4j and an ELK stack https://www.elastic.co/.

Tracking Disinformation in A New Country

America and the UK are original masters at disinformation campaigns, both for their work from second world war onwards, but also for the internal propaganda work so successfully picked up later by e.g. China. Russia, China, Iran are all biggies right now in online disinfo aimed at other countries, but there are also countries whose internal - aimed at their own population - disinformation campaigns have been masterful e.g. Venezuela, or unsubtle but effective, e.g. Philippines. There are other countries where the use of disinformation is just kinda background normal politics, but generally internal and local, e.g. Nigeria. A very subjective top 10 list would be: USA, China, Russia, Iran, UK, Saudi Arabia, Pakistan, India, Venezuela, Philippines.

Things to think about: who

- How is a country involved?
    - Disinformation customer / originator
    - Disinformation target
    - Disinformation producer / factory
- What type of disinformation?
    - Geopolitics / Nation State propaganda: country A to country B/C/etc
    - Politics / propaganda: country A to own population
    - Grifting: individuals to population (usually for money)
    - Power: groups to population (recruiting, actions etc)

Things to think about: what

- Localisation:
    - Local tech use (including social media)
    - Local power structures
    - Local concerns
    - Languages
    - Communication style
    - Local idioms (e.g. “cockroaches”)
- Globalisation
    - Common themes: politics, grifters, 5g, antivax etc