Big Book of Disinformation Response

1 Table of Contents

2 Introduction

This is the big book of disinformation response for the CTI League's disinformation team. We're embedded within the CTI League, and track disinformation using similar tools and techniques to the rest of information security, but there are some things that we do a little differently. Hence this book.

2.1 The CTI League

The CTI League is a community of cyber threat intelligence experts, incident responders and industry experts working to neutralize all cyber threats looking to exploit the Covid19 pandemic. It identifies, analyzes and neutralizes all threats but at this most sensitive time is prioritizing front-line medical resources and critical infrastructure. The League's April 2020 report is https://cti-league.com/2020/04/21/cti-league-inaugural-report/ and its activities are listed in the playbook.

The disinformation team is tasked with finding coordinated inauthentic activities ("disinformation campaigns"), and the objects and people attached to them, and using known ways (some of which we have to invent ourselves) to expose, disrupt or stop their operations.

2.2 Glossary

We all come from different disciplines: words like "campaign" have different meanings to a military, an adtech or a tech person (and if you're all three, you get to fight about definitely with yourself). There are also committees dedicated to defining what words like "disinformation" and "misinformation" mean, and the differences between them.

We ain't got time for that here. This glossary is our latest best effort at definitions for some of the words we use a lot between us, and what we (mostly) think we mean when we say them.

- **Cognitive Security**: The top layer of security, alongside Physical-security and Cyber-security. The art and practice of protecting against hacks that exploit cognitive weaknesses, especially cognitive hacks that are online and/or in large numbers of people. One of the reasons the MisinfoSec crowd started talking about Cognitive Security (including rebranding as the CogSecCollab) in 2020 is a belief that, in order to deal with things like disinformation, we need to focus on the thing we're protecting. That

means working on reducing disinformation, but also on boosting good information when we see it.

- **Misinformation**: false content, where that content could be text, images, video, voice etc. Misinformation does not have to be deliberately generated (e.g. my mother might forget my favorite colour)
- **Disinformation**: deliberate attempt to deceive online. There is usually intent to deceive with disinformation, and the content itself might be true, but in a deceptive context (e.g. fake users, fake groups, mislabelled images, doctored videos etc). Claire Wardle's work on the differences between misinformation and disinformation is still some of the best.
- **Campaign**: Campaigns are long-term efforts to change or confuse populations.
- **Incident**: Incidents are coordinated inauthentic activity that are carried out as part of a campaign. The "coordinated" implies either an instigator of some form with motives (geopolitics, money, ideology, attention, etc.) or some form of collective deliberate behaviour around it, like flooding a hashtag. That activity usually lasts for a short period of time because the narratives, artefacts, and other aspects can be picked up and continued by people who aren't driving an incident - and this is often part of an incident or campaign's goals.
- **Narrative**: Narratives are the "stories" that are being used to change minds, confuse people etc. 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.
- **Artefact**: Artefacts are the objects that you can 'see' connected to a disinformation incident or campaign. They're the text, images, videos, user accounts, groups, hashtags etc that you use to get a picture of an incident or campaign.

Other terms related to this work:

- **Astroturfing**: creating a fake grassroots movement with an obfuscated sponsor or orchestrating group

2.3 Styles and formats

- We use ISO8601 format for dates where possible: yyyy-dd-mm (see https://www.w3.org/QA/Tips/iso-date)
- When referencing specific times related to incidents, explicitly declare the timezone or use UTC

2.4 Other places to look for information

There's a lot to learn about disinformation, misinformation, and how they fit into cognitive security / infosec in general - there's a separate BigBook of Cognitive Security for all that. This BigBook is the practical one.

We've added lists at the end of this document (here), to books and papers about disinformation, to other teams doing this, sources of data, tools etc. And CogSecCollab is also collecting information in its documentation repo, which was used to seed this document.

For all things CTI Disinformation, start at the Team Readme.

2.5 How Disinformation fits into the League

2.5.1 Activities

Reading through the CTI League handbook, the league stresses *"Our members prioritize efforts on helping hospitals and healthcare facilities protect their infrastructures during the pandemic and creating an efficient channel to supply these services". We should do this.*

It lists services as:

1. Neutralize malicious activities in the cyber domain with takedown, triaging, and escalation relevant information for sectors under threats.
2. Prevent attacks by supplying reliable, actionable information (IoCs, vulnerabilities, compromised sensitive information and vulnerabilities alerting).
3. Support the medical sector and other relevant sectors with services such as incident response and technical support.
4. Act as clearinghouse for data, a connection network and a platform for facilitating those connections

There are disinformation equivalents to these:

1. **Neutralise**: This is the disinformation takedown, triage and escalation work listed under disinformation incident response below.
2. **Prevent**: This is work that we could be doing - collating and supplying disinformation IoCs and vulnerabilities to the organisations, especially the health organisations, that we work with. For example, if we identify that a "Reopen $STATE" campaign is attempting to organize another "Operation Gridlock" incident, we can alert state, city, and county officials, as well as any hospitals in the target area.
3. **Support**: We've seen few direct cognitive security attacks on medical facilities so far. We have seen attacks directed at high-profile medical individuals and general attacks. We can assess the possibility of direct attack, and ways to be ready for that. For example, we could prepare resources that could be used in countering campaigns that target COVID-19 field hospitals (such as the Javitz Center field hospital in NYC).

4. **Clearinghouse**: We have connections established, but haven't built ourselves as a clearinghouse yet. We could [Comment1][Comment2]. We could also coordinate this work with those who are focusing on response and countercampaigns (the "elves" who fight the "trolls").

For the neutralisation part, the league lists as examples:

- Infrastructures used by a threat actor that is exploiting the pandemic – malicious command and control server / DDoS servers / domains / IPs / etc.
- Exploiting legitimate services (such as open port in a legitimate website or compromised website used by hackers) and relevant to our stakeholders can be used to deploy attacks

The disinformation equivalents here would include:

- Hashtags, groups, networks, botnets, information routes, etc used by disinformation actor groups to create and run incidents. We can map several of these ahead of time, monitor them for new events forming (e.g. qanon checkins etc), and also file abuse complaints to registrars etc, notify companies hosting botnets and command and control accounts etc.
- Medical events (e.g. vaccination rollouts) that we know will trigger disinformation incidents

For prevention and support, the league lists examples:

- Alerting about vulnerabilities / compromised information and infrastructure to our stakeholders
- Creating a database of malicious indicators of compromise for blocking (via both MISP and GitHub repository)
- Alerting about trends and uneventful events regarding the pandemic in the cyber domain
- Creating a database of hunting queries for alerting systems.
- Create a safe and secure infrastructure for CTI League activities
- Create reports dedicated for the stakeholders and update them about ongoing trends of attack vectors regarding their organizations, such as significant information from underground-based platforms (darknet).

This is more detailed work, but as we track more incidents and become more familiar with the methods and tools used by incident creators, some measure of prevention activities become possible.

2.5.2 Channels and Bots

We have potential inputs, outputs and help across the other CTI league channels, beyond our own channel #4-disinformation.

- #2 channels are useful for finding us the people and places we need to get assistance, to report to (e.g. to find a specific Twitter group representative), to request takedowns etc.

- #3 channels are supplementary input data
- #4 channels are other teams (e.g. darknet) who work alongside us sometimes on the same artefacts
- We could add outputs to #5 channels
- #6 channels could become useful in future.

3 How our team works

3.1 Coming in to help

The main work of the disinformation team is incident tracking and response. Live incidents are listed in theHive, and new ones are flagged in our slack channels as they're added. We have 5 subteams supporting this:

- Incident management
- Tech
- Outreach
- Process and training
- People

Team leads can be reached by pinging @disinfo-leads in #4-disinformation. To get involved, fill out the disinformation survey.

When in doubt, ask a team lead. Otherwise, checking social media to see if a new incident is brewing is a never-ending job.

3.2 Basic OpSec for our team

3.2.1 Key concepts

- Security. It's a process. Tools help you execute the process.
- Compartmentation: separate your personal life from your work life.
- Persona: your spy disguise for research. A fleshed out human being that has details.
- Step 0: Lock your shit down.
- Goal: Impact containment. If you use compartmentation and a persona and everything goes wrong, all that gets compromised is the persona.

3.2.2 Process

OPSEC is a process, not a set of rules or tools. By continually following the process the user should remain in a state of security. The security you get is from following the process, not using tools.

3.2.2.1 Threat modeling for humans

EFF's Surveillance Self-Defense guide has a great introduction to threat modeling. In general, think about your 1-3 biggest threats -- in our case, revealing your real identity -- and consider the following:

1. What am I protecting?

2. From whom?
    a. What are they capable of doing?
    b. What's the worst that can happen to me?
3. How am I protecting myself and my info? (mitigate against them)

Once you've assessed your threat model, it's important to put it into action. Don't just sit there -- do it!

3.2.3 compartmentalization: Engineering to make mistakes difficult.

An important part of operational security is implementing compartmentalization to limit the damage of any one penetration or compromise. compartmentalization is the separation of information, including people and activities, into discrete cells. These cells must have no interaction, access, or knowledge of each other. This is sometimes referred to as **impact containment**.

By compartmenting your operations, the control center over your accounts, and the information available from any single persona source, you are limiting the impact of a compromise. Without proper compartmentalization, attackers are able to leverage information from one compromised account to access another related account. Increasing privileges and traversing across the persona's exposed and interlinked account control centers.

The strength of this compartmentalization is directly proportional to how strong your compartment walls are, and how well you maintain them. This takes discipline. But it isn't impossible.

3.2.4 Foundations: Personal Security

3.2.4.1 (Step 0) Baseline Security

Before you do anything else...

Secure yourself. Harden your personal environment.

- Implement unique, strong passwords everywhere
- Enable multi-factor authentication (2FA or MFA) on everything.
- Lock down privacy settings on your social media.
- Minimise your attack surface and exposure to retaliation if everything goes wrong.

Additional reading:
Security Guidelines for Congressional Campaigns
EFF's Surveillance Self-Defense Guide

3.2.5 Foundation: Work environment

3.2.5.1 Compartmentalization

No matter how good people get at hacking, they still have to obey the rules of physics.

Machines: Don't use your personal computer. Use dedicated equipment.

- At a minimum, use a Virtualbox VM.
- Better: use a separate, dedicated computer.
- Don't trust your brain to be perfect -- configure your computers differently so you have visual cues.
    - Use separate wallpapers and themes
    - Use separate browsers for separate tasks.
    - If you use dark mode on your personal computer or VM, set up light mode on your research computer or VM

Use a VPN: VPNs tunnel your internet traffic to make it look like you're in a different physical location. Use a paid product; if you're not paying a subscription for your VPN, the provider is collecting all of your traffic and selling the data.

If you're not sure which one, try ProtonVPN or Private Internet Access.

3.2.5.2 Cover: Your Persona

Once you've created your compartmented workspace, it's time to create a persona.
You're not trying to beat the NSA; you're trying to avoid being doxxed by trolls on 4chan. While it can be easy to go down a rabbit hole on this, you likely don't need a lot of backstory. With that in mind use a site like fakenamegenerator.com to create a persona.

Your persona should include at least:

- Name
- Email
- Phone number (non-VOIP burner works best if signing up for accounts)
- Account usernames and passwords
- Address
- Birthdate

Keep this info in a text file and leave it on the desktop of your working machine.

3.2.6 Work recipes (if this, then that)

Need to get people to explain the process of what they're doing, so we can build out the relevant recipes

- OSINT Research

Always start with Step 0: Baseline Security

This is intended as a quick and dirty guide to considering your Operational Security (OpSec). Consider this a starter guide or Level 0. There is a baseline for security to protect yourself, your fellow researchers, and the project. Obviously your approach to OpSec is going to depend on your threat model. Given the current context I'm going to skip an in depth discussion of physical security in favor of other topics.

3.2.7 OpSec Appendices [Comment03]

The starting point for building security is to limit the potential impact of a compromise. To contain the damage from a compromise use the principle of compartmentalization. Build a strong secure compartment to use for all your work and ensure their is no taint or contamination from inside the compartment back to you.

3.2.7.1 Threat Modeling

(from Lorenzo Franceschi-Bicchierai's What is Threat Modeling?)
"The first step to online security is figuring out what you're trying to protect, and who you're up against.

To help you figure out your threat model, consider these five questions:

1. What do you want to protect?
2. Who do you want to protect it from?
3. How likely is it that you will need to protect it?
4. How bad are the consequences if you fail?
5. How much trouble are you willing to go through in order to try to prevent those consequences?

By answering those questions, and figuring what solutions and tools you want to adopt based on them, you will come up with a threat model that works for you.

Overestimating your threat can be a problem too: if you start using obscure custom operating systems, virtual machines, or anything else technical when it's really not necessary (or you don't know how to use it), you're probably wasting your time and might be putting yourself at risk. At best, even the most simple tasks might take a while longer; in a worst-case scenario, you might be lulling yourself into a false sense of security with services and hardware that you don't need, while overlooking what actually matters to you and the actual threats you might be facing."

3.2.7.2 Physical Security Basics

- **Cover your webcam** to prevent unauthorized access to your camera.
- Lock and password protect computer
- Enable full disk encryption
- Optional: If you're concerned about unauthorized access to your microphone, you can use a mic block. Here is one example.

3.2.7.3 Passwords

Weak passwords and password recycling are the easiest ways to have your accounts pwned

- Haveibeenpwned: Check if your email account has been compromised in a data breach.
- Most password managers will alert you if your password has appeared in a data breach.

3.2.7.4 Password Managers

Password managers are the easiest way to create, store, and implement secure passwords for all your accounts.
Decision Point: Local or cloud-based password manager.

- Local: more secure, less efficient, harder to maintain, easier to lose everything if you forget to back up or lose access to your local version
- Cloud-based: easier to use, accessible anywhere, more efficient, less secure

Some options:

- 1password (cloud-based)
- LastPass (cloud-based)

- Dashlane (cloud-based)
- KeepassXC (local)

3.2.7.5 Two-Factor Authentication (2FA)

Two-Factor Authentication requires the user to provide an additional form of verification beyond just their password (Something you have + something you know). After having a strong unique password for each account, adding 2FA to an account is the highest leverage way to secure your account against unauthorized access.

- Two-Factor Authentication Handout from the EFF
- Twofactorauth.org: List of websites and whether or not they support 2FA.

Decision Point: Method for 2FA

- Text message (SMS): Easiest to get users to adopt, least secure, especially in our context. If you use it, best to use a burner VOIP number.
- Soft token (App-based): More secure than SMS. Examples include Google Authenticator and Authy.
- Hard token (Physical device): Most secure, harder to implement. Examples include Yubikey.

3.2.7.6 Using a VPN

A VPN is a program that routes all of your internet traffic through a different IP Address (like a tunnel). A VPN is one of the most effective ways to maintain anonymity online. Since VPN's basically route all your traffic like an ISP would, be sure you trust the provider. This is one of those things you should pay for, because if you're not paying for the product, you are the product. The VPN market is a racket; the review sites are a part of that. I've found thatoneprivacysite's reviews to be useful.

Here are some VPN options I've found helpful:

- ProtonVPN, by the same folks that make Protonmail
- Private Internet Access

Check that you're VPN is working properly by going to ipleak.net

Decision point: VPN on your network, on your device, or both

- On the network:
    - Pro: Filters all traffic from all devices on your network, not just web traffic or one device. If you lose VPN connection you can kill all internet access so nothing gets through without going through the VPN
    - Con: Longer and more complex setup and you need a dedicated device
- On your device:

    - Pro: Quicker and easier to get set up. Doesn't require any extra equipment.
    - Con: Only filters traffic from your one device and if it fails you may not realize immediately (unless it has a reliable killswitch). Also data your computer sends back to services on startup may get through before the VPN kicks in.

3.2.7.7 Web Browsers and Extensions

Decision Point: Which browser to use for general investigations
My browser of choice: Firefox
Essential Extensions

- Install Firefox Multi-Account Containers lets you separate your work, shopping or personal browsing without having to clear your history, log in and out, or use multiple browsers. Container tabs are like normal tabs except that the sites you visit will have access to a separate slice of the browser's storage. This means your site preferences, logged in sessions, and advertising tracking data won't carry over to the new container. Likewise, any browsing you do within the new container will not affect your logged in sessions, or tracking data of your other containers.
- Install Privacy Badger a browser add-on from the EFF that "stops advertisers and other third-party trackers from secretly tracking where you go and what pages you look at on the web.
- Install uBlock Origin, a wide-spectrum content blocker.
- Install HTTPS Everywhere, a browser extension from the EFF that encrypts your communications with many major websites, making your browsing more secure.

3.2.7.8 Burner Email and Phone numbers (pseudonymous identities)

In the process of doing investigations, you will likely find yourself in a position where you want to create burner accounts that allow you create pseudonymous personae. When possible, I create a full identity with name, email address, VOIP phone and text as well.

- Sudo: In terms of an easy to use pseudonymous identity, I've found that sudo is a great, easy to use option. It is a paid service, so that can be a barrier, but it allows you to create a personae and associate and isolate email, phone calls, text, web browsing and payment for each persona.

3.2.7.8.1 Burner Emails

Depending on your needs you may wish to create anonymous/pseudonymous emails. These are disposable temporary email addresses you can use. Many of these will get flagged by social media services as suspicious, so it's good to know about different options.

- 33mail Free option that might get flagged
- Protonmail: Free end-to-end encrypted email

- Gmail: quick and easy commercial option that will pass muster for most services. May have issue with this if you try to sign up for a bunch with the same phone number (which you shouldn't do anyway)

3.2.7.8.2 Burner Phone and phone numbers

There are tons of ways to get a free VOIP account. One challenge with VOIP numbers is that some services you'll want to use require a real phone number and won't accept VOIP for account registration.

- Free VOIP: Google Voice. You'll obviously need an associated Google and account and getting it requires providing a real phone number (major downside).
- Paid VOIP: Burner, Hushed, CoverMe
- Burner phones: Lots of different options including Tracfone where you can get a cheap phone and swap the SIM when needed.

3.2.7.9 Secure Communications

Use End-to-End Encryption (E2EE) wherever possible. E2EE is a system of communication where all data is encrypted in transit and at rest, meaning no one (including employees at the company) has access to the data except the communicating users. This is the closest you're going to get to a completely private and secure way to communicate and store data.

3.2.7.9.1 Secure Messaging

End-to-End Encrypted messaging generally requires both users to be on the same service. This often means that the best service is the one with the most people you're trying to communicate with. Here are a few options:

- Signal is great and the How to Use Signal on iOS from the EFF is helpful. Popular among infosec, privacy enthusiasts, and journalists. One downside is that you have to tie the account to a real (non-VOIP) phone number.
- Whatsapp: Most popular E2EE messaging app. Built on the same encryption protocol as Signal. Major downside: owned by Facebook.
- iMessage: Incredibly popular. Only available to Apple users. E2EE breaks down depending on how you configure its relationship to iCloud for backing up messages.
- Others: Wire, Wickr, etc

3.2.7.9.2 Secure Email

End-to-End Encrypted email services:

- Protonmail
- Tutanota

3.2.7.9.3 Secure Ephemeral Communications:

- Firefox Send uses end-to-end encryption to keep your data secure from the moment you share to the moment your file is opened. It also offers security controls that you can set. You can choose when your file link expires, the number of downloads, and whether to add an optional password for an extra layer of security.
- CloakMy: quick, convenient and secure way to share sensitive information. Just copy your message in the box, set the recipient and your password (if you want to protect your message) and send it. The recipient will receive a secure link. If you select Auto Destruct as an expiration setting (by default), once the link is opened the message will be deleted. The message will be encrypted with a randomly generated key + your password if you chose one.

3.2.7.10 Social Engineering and Phishing

Phishing happens to everyone and it sucks. Here are a few ways to avoid getting phished.

- Urlscan.io allows even inexperienced users to investigate possibly malicious pages, such as phishing attempts or pages impersonating known brands.

A few other things to consider (which I hope we can expand upon later)

- Turn off location services on everything possible
- Locking down the setting on your social media accounts
- Removing yourself from people search sites (in case you get doxxed)
- Remove metadata from your photos before you post them
- 'This person does not exist' generates very convincing faces, again using machine learning. Reload the page to see another image. As the name suggests, these are not real people - the faces are generated entirely automatically. You can see artifacts, especially in the teeth, but this is still very close to perfect (and of course great for creating fake users).

3.3 Mental Health

Disinformation includes difficult material - it's often designed to increase emotions like fear, hatred, disgust, to form in-groups and out-groups with hate speech and images that can be difficult to view, especially if they're of a group you're part of or feel strongly about. Even those of us who've been handling this material for years still get affected (that's the point of it), so we all need to look after ourselves.

Some basics:

- Pace yourself if you're going through difficult material.

    - Take regular breaks. Don't spend more than an hour at a time reading through material.
    - If you can, arrange to be interrupted. It's easy to get into a spiral with difficult material, and find yourself hours later still digging through it. Having an alarm, or a scheduled call from a friend, or the dog pestering you for its walk etc at the end of a session can stop this happening
    - If you can, go through material with a 'buddy' - pair up with someone online, preferably with a video or audio channel, and talk through what you're doing with them.
    - Chocolate helps. We have no idea why.
    - If you start feeling wibbley, stop. There is no shame in this. Nobody in this team will ever judge you for taking a day, a week, two months off to look after yourself, or even shifting focus forever. Your mental health is important, and we will still be here when you're ready.
- If you can avoid touching or reading material, do so. That means that, where we can, we automate. If we have 50 copies of the same image, we only need to view one copy, and if it's a difficult image, not everyone on the team needs to see it. [Comment04]
    - If you have to share images / text in channels, put them in threads below content warnings, so people can choose whether to view them or not.
    - Automate feeds: if we have 50 copies of a message or image, only show 1 copy to the humans.
- Make disinformation something you "go to". Right now, we're surrounded by "the infodemic". Friends are talking about it, feeds are everywhere, your great uncle is probably selling you the latest conspiracy theory. We're also seeing most people in our lives online. Your life needs to include puppies and kittens, not being swamped by batshit crazy disinformation…(See Basic OpSec for our Team section above)
    - Don't use your main social media accounts to follow disinformation. You don't need more of that in your life. Pull the data you need using APIs; set up dedicated accounts to do the follows; ask the team if someone's already following the accounts or groups you need data on.
    - Incognito mode. Nobody needs their ad feed full of Qanon t-shirts and bleach cures.
    - We won't always be passive, so having some active accounts could be useful too...

4 Disinformation

4.1 Good introductions to disinformation

Things to show your mum

- The War on Pineapple: Understanding Foreign Interference in 5 Steps
- Bad News Game https://www.getbadnews.com/#intro
- The Dark(er) Side of Media: Crash Course Media Literacy #10
- Web Literacy for Student Fact-Checkers – Simple Book Production

4.2 Disinformation Layers

Disinformation pyramid

As we explore and analyze the information sphere, analysts have techniques that are employed to understand disinformation operations - and they're classified using similar frameworks to those we use to classify our understanding of other types of threats and incidents.

- Campaigns: are long-term disinformation operations. They're focussed around a theme, like specific geopolitics (e.g. "make everyone like china" or "Ukraine is really Russia"), and are often nation-state-funded, but might also be from interest groups (e.g. far-right-wing, antivaxxers etc).
- Incidents: these are the short term, cyclic things we track. They're coordinated sets of activities that happen over a defined timespan that usually indicates some form of team or individuals driving them. Incidents have things with defined parameters like TTPs that we can share, threat actors, and other objects that you'd recognise from TI, but also including context and narratives.
- Narratives: are the stories that we tell about ourselves and the world. They're stories about who we are, who we do and don't belong to, what's happening, what's true (e.g.

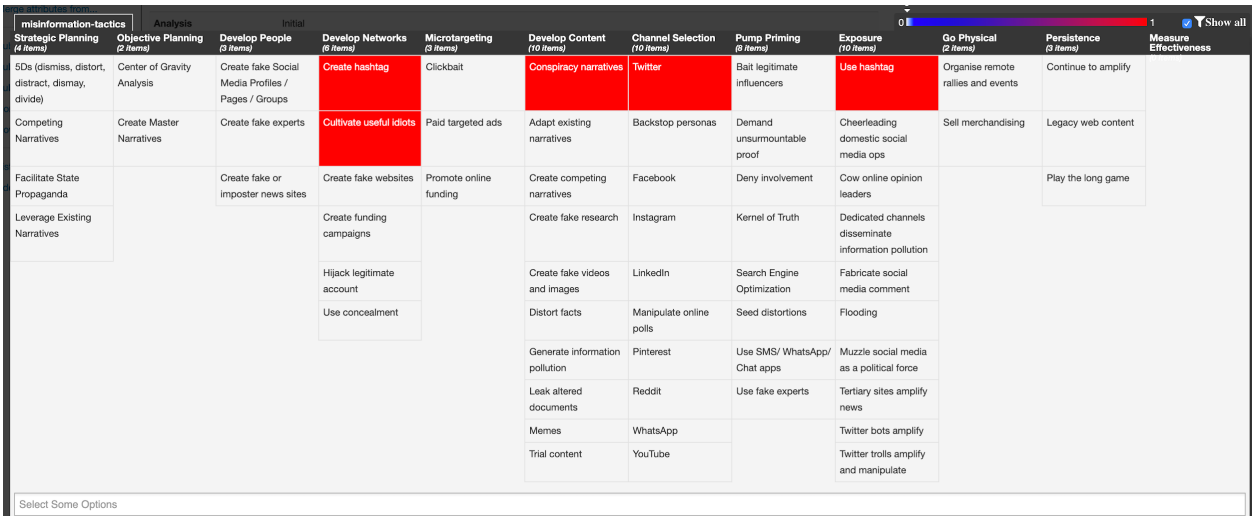
Covid19 was caused by 5G masts). Tagging information with defined narratives make it easier for us as analysts to follow the flow of information across the internet and beyond.

- Artefacts: Incidents and Narratives show up online as artefacts: the text, images, videos, user accounts, groups, websites etc and links between them all that we collect and use to understand what's happening.

So what looks to outside observers like analysts simply hunting down a hashtag or a URL, describing a narrative, or trying to understand the things that link to it is so much more; it's really a part of creating an inventory of the discrete elements of each incident, or the objects used by a disinformation team or campaign, so we can a) share a summary of what we think is happening, and b) disrupt both those component parts, the TTPs behind them, and the incidents and campaigns they support.

This is a lot of text. And we're realising that there's a lot of stuff we haven't explained. So we're writing it down. And making stuff clearer and cleaner to use as we test and explain it. This document is those explanations.

4.3 Disinformation TTPs



AMITT, as seen in MISP

We're using the AMITT framework to break each disinformation incident down into its component TTPs and TTP-level counters.

4.4 Covid19 Disinformation outside the USA

Who does this? I would put America and the UK at the top of my original masters list (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 https://www.cbc.ca/radio/ideas/how-hollywood-became-the-unofficial-propaganda-arm-of-the-u-s-military-1.5560575). 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 (Venezuela) or unsubtle but effective (Philippines). There are other countries where the use of disinformation is just kinda background normal politics, but generally internal and local (e.g. Nigeria). My top 10 list? USA, China, Russia, Iran, UK, Saudi Arabia, Pakistan, India, Venezuela, Philippines. [Comment05]

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 [Comment06]

Places to look for non-USA disinformation:

- Disinformation repositories
    - https://euvsdisinfo.eu/disinformation-cases/ - Russia disinfo on EU
    - https://medium.com/dfrlab - world disinfo
    - https://comprop.oii.ox.ac.uk/ - nationstate actors
        - Specifically The Global Disinformation Order and case studies

      https://www.newsguardtech.com/covid-19-resources/ - c19 domains for several countries
- Hive cases, MISP events etc
    - E.g. reopen starting in Australia, moving to Canada etc

References

- India:

    - https://theasiadialogue.com/2020/04/06/india-covid-19-misinformation-and-the-downside-of-social-media/ - whatsapp, fake cures, SM responsible for curation, strong messaging from Modi
    - https://www.weforum.org/agenda/2020/04/indian-scientists-covid19-false-infomation-coronavirus/
- Italy:
    - https://medium.com/dfrlab/italian-mp-amplifies-debunked-covid-19-conspiracy-theories-on-the-floor-of-parliament-fa0a88999142
- China:
    - https://www.aljazeera.com/news/2020/03/china-coronavirus-propaganda-push-ties-worsen-200325085419818.html - hero story
- Africa:
    - https://www.bbc.com/news/world-africa-51710617
    - https://africanarguments.org/2020/03/26/the-other-covid-19-pandemic-fake-news/
    - https://www.facebook.com/watch/NCDCgov/ - countering
- Venezuela:
    - https://www.atlanticcouncil.org/blogs/new-atlanticist/the-coronavirus-infodemic-in-latin-america-will-cost-lives/
- Ecuador:
    - https://www.miamiherald.com/news/local/news-columns-blogs/andres-oppenheimer/article241929726.html - targetted by bot farms from neibouring countries
    - https://www.chequeado.com/latamcoronavirus/ - Latam countering (case lists, in Spanish)

4.5 Where to direct non-Covid19 disinformation

It's almost certain that in the course of looking for Covid-19 related disinformation, we're going to find disinformation on other topics. While our mandate is specifically Covid-19 related, there are other, area-specific organizations to which we can report disinformation.

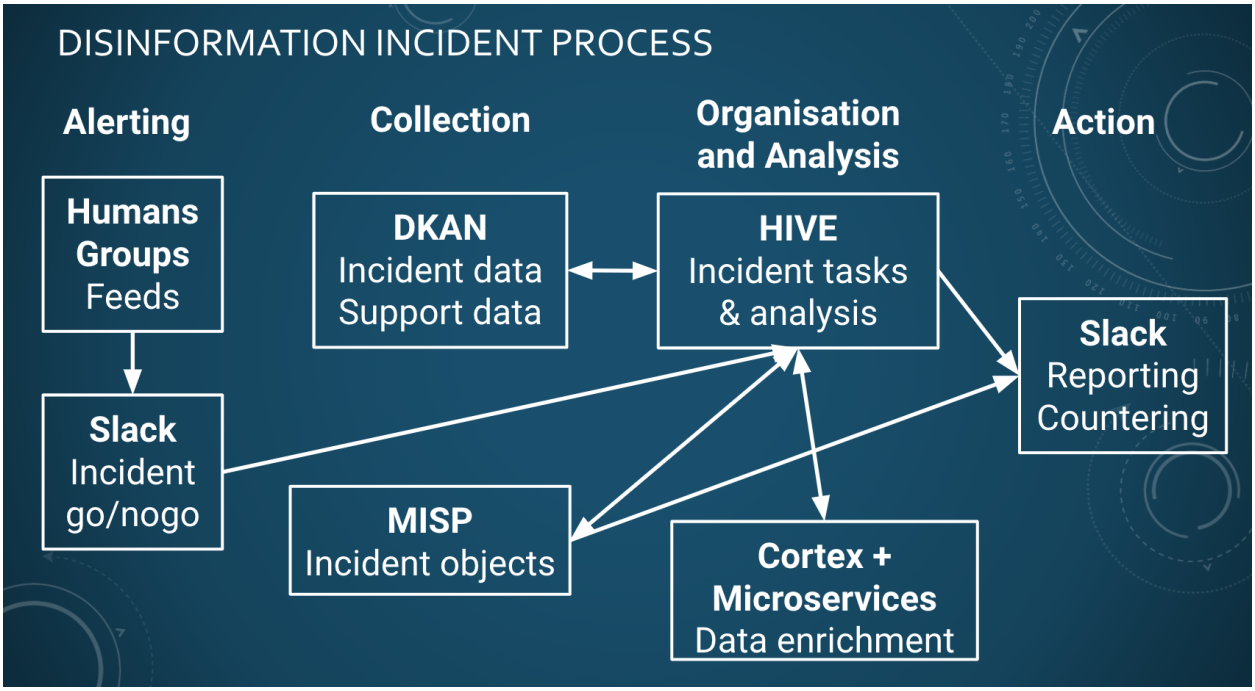
- U.S. Election security: Open a Slack DM to anybody working at Department of Homeland Security, specifically Spencer Wood or Stacey Wright
- Right-wing extremism/hate speech: Southern Poverty Law Center, at https://www.splcenter.org/reporthate
- Voter suppression attempts:
    - On social media, report the post to the platform using their reporting mechanisms
    - You can also report the issue to the U.S. Department of Justice using https://civilrights.justice.gov/report/
- Anti-GLBTQ+: Gay and Lesbian Alliance Against Defamation, at https://www.glaad.org/form/report-media-defamation

5 CTI League Incident Workflow

Disinformation workflows include:

- Tracking an incident
- Tracking narrative flows across incidents
- Adding and maintaining supporting disinformation data

5.1 Incident workflow



Incident dataflow

The main workflow in the disinformation team is tracking an incident. We've broken this into 4 stages: alerting, collection, analysis, and action.

5.2 Workflow instructions

Starting a disinformation incident:

A new Covid19-related rumour has started online. You've seen it yourself, someone has sent you an example of it, you've seen another group tracking it - there are a bunch of ways to spot something new happening. Now what? **NB each of these steps can be by different people**

1. Tell people
    a. Put a message in slack #4-disinformation, with the artefact you found and a short description.
        - Start with "NEW RUMOR" so we will be able to track them

        - Any supporting information or links (under that rumor) should be posted in a thread off that initial NEW RUMOR post
        - This will make documenting and adding objects and observables to the incident and analysis log easier to track, and also keep everything a little more tidy
2. Decide whether to start an incident [Comment07]
    a. Do a quick check that it's a rumour. One sighting doesn't make an incident. 15 copies of the same message on Twitter, or 3 friends sending you the same strange DM, and you're probably onto something.
3. If it's significant, start an incident
    a. Give it a name. Names help.
    b. Add a row to the incidents spreadsheet
    c. Create a folder in the googledrive INCIDENTS folder for notes and anything that won't fit into the DKAN
    d. Start adding data to the DKAN (learn more about DKAN here)
4. Investigate the rumor
    a. Look for related artefacts, accounts, urls, narratives etc
5. Investigate ways to close down the rumor / repeater sites etc.
6. Report on the rumor
    a. Add an incident to the MISP instance for this rumor (learn more about MISP here)
        - The incident must include some relevant observables such as a Tweet, social media username or URL.
    b. Write and send notes/reports to the people who can respond
    c. Close down the rumor and move onto the next one (there's always a next one)

Help with a disinformation incident

1. The master document for what we're doing on incidents is the incidents spreadsheet. Look at the status column - the priority is live incidents, then monitor long-term, then "keep an eye on it" (the potential 'zombie' incidents that are probably dead but might restart)
2. Check back in the slack channel, and in the incident README in the googledrive INCIDENTS folder to see what's been done with this incident recently. As we get things together, we'll probably have incident-specific tasks in the github issues list, but we're still working on that.
3. Find articles and artifacts, investigate the ones we have, put results into the slack channel for harvesting by the bots, and/or discussion with the team.
4. If you spot something significant (new objects tied to the incident etc, new things of interest), update the incident README.

5.3 Alerting [Comment08][Comment09][Comment10]

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

- Alerts from disinformation team members
- The covid19activation slack group (the Tedx team feed)
- The covid19disinformation slack group (the Atlantic Council team feed)
- CTI League Phishing inputs - maybe not so much; lots false positives
- Phone honeypots
- disinfo@ctileague.org - reporting hotline
- Feeds potential from other groups - e.g. peacetechlabs have offered a feed
- Mitre covid19 feed [Comment11]- might be in wrong direction; needs to be symmetric
- Sniff EuVsDisfo - is slow (narrative based) – [Name Redacted]'s dataset/ data stream list
- Sniff hamilton68 [Comment12] dashboard for themes
- Sniff botnet feeds for themes
- Set up reporting from Facebook, twitter etc
- Ask Facebook for feeds from them
- New data coming into the DKAN

At the moment, all the team's feeds are manual; team members check other slack channels etc, or CTI League members post alerts in the 4-disinformation slack channel. 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 #4-disinformation 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. An incident needs to be within the team's scope, and large enough to be worth expending team effort on.

- CTI League is Covid19. Do we just cover Covid19? No - can include politics. Don't care about aliens though.
- Anybody in the disinformation team can start an incident, but the group decides what it reports on.

When we see an alert, we have some questions:

- Is this an incident, e.g. is it a large coordinated disinformation incident, or an isolated piece / few pieces of disinformation?
- Is this disinformation suitable for processing by the disinformation team (e.g. 419 scams might be better handled by the Phishing team, but might also contain information about incidents that we should check out too)?
- Is this disinformation already being handled by platform teams or other specialist teams (we might want to check in with them just in case, for instance referring to #3-medical-sector-supporting if it is healthcare-related, or issuing a #4-takedown-request because of a finding)?
- Is this incident something that we should track?

“Is this incident something that we should track?”, e.g. how do we choose which incidents to track?

- We don’t track incidents for fun or interest. We track the ones that we have a reasonable chance of doing something useful about - whether that’s raising the alarm to groups or organisations that can respond to the incident, asking them to take specific actions (like taking down a disinformation account or site), or taking actions ourselves (like amplifying counternarratives).
- We also track and counter incidents that we believe give us the best chance of a positive effect, and in the Covid19 deployment, ideally one that impacts health.
- Yes health. We prioritise that over other incidents, although we will include disinformation around current events where they impact populations.

5.4 Hive lists for starting an incident

When you create a disinformation incident in HIVE:

- Create a new case. Use case template “Influence Operation Incident”.
- Name the incident (use this name in all the tools)
- Create an event in MISP for the incident:
    - https://misp.cogsec-collab.org
- List the risks and potential real-world consequences from this incident
- List any time bounds on the incident, e.g. are there events that it’s gearing towards etc
- List any geographical or demographic targets in this incident
- Create a DKAN directory for the incident

MISP list for starting an incident

- List actors and other objects that are important in this incident - we’re using a combination of STIX and DFRlab’s Disinformation Dichotomies standard for this. Add these to the Clean MISP
- List the tactics and techniques that are being used in the incident - we’re using AMITT for this (the version that comes as standard in MISP). Add these to the MISP event.

5.5 Organisation

Documenting analysis:

- We have DKAN and MISP, but also useful to have a google folder for each incident for other things that don’t fit into those, like research notes
- Classifications: if it’s openly available online, then it’s okay to put through e.g. Tableau; if it’s come though internal routes (e.g. SMS), then keep it off public internet (don’t share).
- looking for related artifacts, urls, narratives etc

Who we communicate to: [Comment13]

- Report when something significant happens - e.g. see this main effort for this new line
- Report on time period… if big, a daily report; if smaller a weekly report

- No report goes out without at least 2 people beyond the editor going over it
- End users are also watching the MISP

Who makes decisions:

- Depends on decisions
- Need a board - vote via slack; person calling for vote does @channel to board, or emails them [Comment14]
- Who can add an incident? Anyone can start an incident.
- Who can release a report -
- Who can talk to customer/ victim? Needs to be agreed on

5.6 Collection

DKAN holds data we don't want to lose, and data that's raw and large: it's the in-tray
Data inputs for DKAN

- Potential starts of incidents
    - Feeds from messenger dms (about 30) - on personal facebook/messenger
    - Data in covid19disinfo team slack repository channel
    - Data in covid19activation disinfo-watch channel
    - <sms honeypots>
    - <emails to disinfo email address>
    - <feeds from other groups>
- Analysis datasets
    - Covid5g twitter data (5 directories) - on pc
- Supporting datasets
    - Narrative lists (CMU etc)
    - Narrative descriptions (EuVsDisinfo etc)

MISP hold objects of interest and the relationships between them, so we can quickly look up things we've seen before etc

Data we build up in MISP

- Incidents
- Narratives
- Actors
- Urls

5.7 Action

What we want to do with an incident is disrupt it as much as possible. If we can stop it completely, that's a big win, but generally, we're after disruption. CogSecCollab has a long-list (here: https://github.com/cogsec-collaborative/amitt_counters/blob/master/tactic_counts.md) of the things we can do to disrupt incidents at different stages of the disinformation killchain (https://github.com/cogsec-collaborative/amitt_framework - that, and DFRlab's object labels

https://github.com/DFRLab/Dichotomies-of-Disinformation are what we're using in the MISP reporting), but frankly it's still messy so at this stage it's better to put our hacker hats on and think "which artefacts (observable objects) do we have in this incident, and what can we do to make them less effective?"

Examples: are there URLs pushing out covid5g disinfo? Are there social media accounts and groups pushing out covid5g disinfo? If we gather evidence on these, we can get that to the social media companies. Are there botnets involved (yes, yes, I said the b word, but they're part of this too)? Can report those too. Etc etc (and I suspect many of you have etcs CogSecCollab didn't think of when they created that counters repo).

This is the practical part of incident handling. We track an incident until the underlying incident stops or slows significantly (or the event it's building up to has passed), or until we've done as much as we believe we can to counter it, or know that there are other teams dealing with it.

Disinformation counters are much more than "remove the botnets" and "educate people". For most incidents, there are a variety of things that can be done about the incident, its creators, the objects used in it, and the tactics and techniques used. We've collected a few (well, a couple of hundred) suggestions for technique-level counters at https://github.com/cogsec-collaborative/amitt_counters - we're expecting to uncover a bunch more as more infosec people do disinformation.

5.8 Managing an incident response [Comment15]

An individual can track an incident on their own - open up some notebooks, fire up the coffeemakers and mainline chocolate for a couple of days. That's - not sustainable over time and large numbers of incidents, any more than it is for other infosec incidents.

The short instructions for managing a response are in the team readme. This is some of the thinking around them:

We haven't worked out exactly how to fit cognitive security / disinformation response into a SOC yet, but here's where we are at the moment on starting an incident:

- Incidents need names. Yes, yes, I know that's a slippery slope that ends up in a cute mascot and a dedicated website, but a name makes it easy to quickly identify what you're working on, find the right folder to put things into etc.
    - Action: Make up a name: make it short but descriptive - you're going to be typing it a lot, but you also want to remember what it was about a week later.
- The team needs to know you started an incident - both the team who are around at the time (and can help look for artifacts, add their specialist skills etc), team members who are coming in looking for things to do later, and leads who are trying to balance the load on the team overall. Best way to do this is to add a note to the team chat and an entry in the team log. [Comment16]

    - Action: add a note to the team slack channel, naming the incident and asking for help with it (if needed). If you have a starting artefact, add that too. Adding the word “NEW” will make it easier to find by people looking in on the channel later.
    - Action: add an entry in the team log, saying you’re starting an incident response. At the moment, this is the incidents spreadsheet - this is likely to shift to adding a case to an incident tracking tool like TheHive. [Comment17]
- You, and the team, are going to start producing notes and artifacts as you track through the incident. Create a place to put them, that’s accessible to the team
    - Action: create a space to put images, artifacts etc in. At the moment, that’s creating a folder for the incident under the INCIDENTS googlefolder [Comment18] - this is likely to shift to directly uploading to a tool like TheHive or MISP.
    - Action: create a notes log for the incident. At the moment, that’s a README file in the incident googlefolder - this is likely to stay the same for the moment. In the log, write a short description of the incident, and how you started tracking it (e.g. what the first artefact(s) you saw were.

Here’s where we are on managing investigating the incident:

- You, and the team, are going to investigate the incident
    - Action: Look for related artefacts, accounts, urls, narratives etc
    - Action: add artefacts to the space you set up for collecting images, artefacts etc. You’ll find it helpful if you number the images, because they’re difficult to reference otherwise (aka “the yellow poster again” isn’t as specific as “image001_yellowposter”)
    - Action: keep the flow of investigation moving - keep a list of actions related to the artefacts, and/or direct the team to areas that need further research
- You’ll also need to translate that into an incident description that can go out as an alert to other teams, and be used to look for potential counters
    - Action: add incident to alert tools. We’re using MISP here, so adding a MISP object for the incident, and attaching the objects important to it is appropriate here.
    - Action: map artefacts seen to tactics and techniques. MISP includes AMITT - you can use the ATT&CK navigator to click on all the tactics and techniques you can see in this incident.
    - Action: Investigate ways to close down the rumor / repeater sites etc. We’re working on tools for this too, but for now it’s discuss this with the team, and check the lists below.
- Oh, and yes, you get to be scribe for the team too, making sure you keep a record of the investigation:
    - Action: keep the incident log updated with any significant findings, notes, things to do etc.

And here’s where we are on managing responding to the incident:

- You need to get information about the incident out to other teams that could do something about it: [Comment19]
    - You've already added an incident to MISP; make sure it's ready to go (question: is there something we need to do to get it out on the feeds?). [Comment20]
    - Write and send notes/reports to the people who can respond
- If you found ways to respond, decide what to do, and check whether you did it
    - If the team found ways it could respond - triage them. Find ways to do the ones you can.
    - Also check on the things you were going to do. Was something done? Chase it up.
- And finally, know when to stop.
    - If you've done as much as you sensibly can, close down the rumor and move onto the next one (there's always a next one).

There are always more incidents, although we're often lucky enough to have a few days without anything major going on. Every morning, one of the leads (often [Name Redacted]) looks through the list of incidents and decides which ones should continue to be 'live', which we should move to just keeping an eye on, or keep a longer-term watch on in case they flare up again, and which we can close down as unlikely to be active again.

6 CTI League Other Workflows

6.1 Narrative workflows

Narrative: 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. [Comment21]

But there are a lot of them. Hence the mindmap, which starts to group narratives into hierarchies, making them easier to read and manage.

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.

So 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.

7 Collecting Incident Data

7.1 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").

7.2 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.

7.2.1 covid19-related disinformation data feeds

Narratives

- Wikipedia list of Covid19 rumours
- WHO Covid19 myths list - narratives
- EuVsDisinfo database https://euvsdisinfo.eu/disinformation-cases/
- Ryerson Claimwatch dashboard
- CMU IDEAS Center list of Covid19 disinformation narratives (click dates)
- Indiana Hoaxy (twitter, articles)

Data

- Botsentinel: lists “trollbots” (bot-like and troll-like accounts) and the themes they’re promoting https://botsentinel.com/ (not just Covid19)
- Hamilton68 - live feed from accounts attributable to Russia or China (may or might not contain propaganda; useful for seeing current themes). Public version is live feeds from official Russian sites (embassies, RT etc), not trolls. Academics can ask for a more detailed feed. https://securingdemocracy.gmfus.org/hamilton-dashboard/ (not just Covid19)
- Ryerson University covid19 misinformation portal: https://covid19misinfo.org/
    - Botswatch dashboard https://covid19misinfo.org/botswatch/
- Uni Arkansas COSMOS Covid19 list http://cosmos.ualr.edu/misinformation
- Indiana University OSOME Decahose
- Facebook datafeed: Enabling study of the public conversation in a time of crisis

Domains

- Coronavirus Misinformation Tracking Center – NewsGuard

7.2.2 Covid19-related counter-disinformation feeds

- Ryerson University covid19 misinformation portal: https://covid19misinfo.org/
- Snopes: https://www.snopes.com/
- WHO COVID-19 site: https://www.who.int/health-topics/coronavirus
- WHO information network for epidemics https://www.who.int/teams/risk-communication
- Coronavirus Tech Handbook https://coronavirustechhandbook.com/misinformation [Comment22]
- Experts list https://twitter.com/jeffjarvis/status/1254038157244456961
- Maryland Covid19 rumour control https://govstatus.egov.com/md-coronavirus-rumor-control

7.2.3 Covid19 general data feeds

- https://crisisnlp.qcri.org/covid19 - GeoCov19 dataset of covid19 tweets (up to about 3 weeks ago; still collecting)

7.2.4 General disinformation datasets

- Twitter IO archive: covers several countries up to a few months ago. Good for getting a sense of the size and ‘feel’ of typical nationstate twtter posts/ networks etc. https://transparency.twitter.com/en/information-operations.html
- Facebook ad library: contains all active ads that a page is running on Facebook products https://www.facebook.com/ads/library/ (About the Ad Library)

7.3 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 some on the artefacts you found, but here are some of the ways.

7.3.1 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 https://github.com/r0zetta/twitter_gather

We have code based on an early version of Andy Patel's twitter_gather code 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.

7.3.2 Facebook data

The Facebook API is horrible. Most everyone tracking social media uses a third party like CrowdTangle (which isn't free) or scrapes for the data they want.

7.3.3 Reddit

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

7.3.4 Multi-platform tools

Reaper collects from a set of social media feeds. Trying that out.

Access tokens:

- Facebook: look at list in https://developers.facebook.com/docs/facebook-login/access-tokens/ - then used https://developers.facebook.com/tools/explorer/ to check token worked before putting into reaper.
    - “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
-

8 Handling Artefacts

Artefacts are the things we can see online - they're what we track and use to understand what's happening in an incident, how everything in it fits together, and what we can usefully pass on as information about it at the incident level, or usefully do to influence it. [Comment23] The artefacts that we see most often include:

- Tweets
- Twitter accounts
- Facebook groups
- Domains - websites
- Hashtags
- Images
- Videos
- Audio fragments (e.g. voice messages)
- yas

The next layer up includes:

- Narratives
- Botnets

The basic questions: What is this thing. How is it impacting the things we care about? Are there other teams doing something about it? What can we do about it? How much impact can we make in the things we care about, for the resources we need to expend?

8.1 Handling Domains (URLs)

8.1.1 Chasing a URL

So you've got a URL. Now what? Well, you probably want to know about the URL - who created it, when, what's it connected to etc.

Check for company

- Is anyone else tracking this url? Check reddit etc. - you might save yourself time if other groups have already tracked lists, social media etc.

All the Google Dorks! (h/t to [Name Redacted])

Using Google Dorks to Check Primary sources (from Henk van Ess's Finding patient zero)

Websites as primary sources: This is useful when your searches within specific sites or urls are coming up empty
Step 1: Look at the failing link

- Ex. https://www.sec.gov/litigation/apdocuments/3-17405-event-11.pdf
- Pull out just the domain name and Top Level Domain (Ex. sec.gov)

Step 2: Use “site:”

- Go to a generic search engine.
- Start with the query (“Dutch police”) and end with “site:” followed directly with the URL (no spaces).
- Ex. “Dutch police” site:sec.gov

Step 3: Adapt the “primary source formula” to your needs

- Include specific folders (Ex. “Dutch police” site:sec.gov/public)
- Predict folders you think might be there

Following the trail of Documents

Step 1: Establish the document type

- Is it a doc | pdf | xls | txt | ps | rtf | odt | sxw | psw | ppt | pps | xml file?
- Use filetype: and the type of file with no spaces (Ex. “filetype:pdf”)

Step 2: Include a phrase you’d like to search with in the document (could include a date)

- Ex. You’re searching for an invitation to an event from May 13, 2014, event. (Be sure to search for both the cardinal and ordinal forms, May 13 and May 13th.)

Step 3: Who is involved?

- Do you know the creator/host and it’s website?
- Ex. The organizer is “Friends of Science” and its website is friendsofscience.org.

When you combine all three steps, the query in Google will be:
“May 13th, 2014” filetype:pdf site:friendsofscience.org

Filtering social media for primary sources

YouTube

YouTube’s search tool has a problem: it won’t let you filter for videos that are older than one year. To solve this,

- In a Google search include the keywords and site:youtube.com
- manually enter the preferred date into a Google.com search by using the “Tools” menu on the far right
- Then select “Any time” and “Custom Range.”

Process for investigating the authenticity of a website :

Web searching a domain: Since we want to find out what other sites are saying about the site while excluding what the site says about itself, we use a special search syntax that excludes pages from the target site

- Search syntax is website -site:website
- (Ex. baltimoregazette.com -site:baltimoregazette.com)
- Scan the set of results looking for sites we trust

Finding out who runs a site with WHOIS :

- Enter the domain name into the WHOIS Domain Tools

- Note who the domain was registered to
    - Unfortunately, WHOIS blockers have dramatically reduced the value of WHOIS searches, so you may only find a proxy.
- Note when the domain was registered

Use a backlink checker like ahrefs or smallseotools that allows you to see all websites that link to a particular site

Look up the url

- Builtwith:
    - look on builtwith.com - if you're lucky that will tell you when and who
    - It will also tell you which sites have the same tags as this site: this helps you find connected sites
    - Use CSC code run_builtwith.ipynb - same thing, but gives you json and a dataframe of those connected sites

Look at the URL contents

- Are there phrases you can use in a googlesearch, to find related objects? Run the search that allows repeated results, to see identical pages. About and terms pages are usually good places to look for these.
- Use CSC code googlesearch_for_terms.ipynb to search for terms/ pages.
- Are there people connected to the site? Start searching for them
- Are there companies?

Think about geography etc

- Look at the title and url of the site. Do they have elements that might be repeated? E.g. if you have xxxmichigan.com, check for the same pattern with other states' names, e.g. xxxwisconsin.com. Astroturfers try to cover an area, whether it's geographical or demographic, and if they're doing it for money, they'll usually have multiple sites.

Look for links

- Check social media - are there references to the URL, or groups / pages / accounts with the same name?
- If there are references to the URL, are there common hashtags, phrases or people in common you can use to search for more sites?

Examples:

- http://overcognition.com/2020/01/22/data-safari-rough-notes-pink-slime-network/

8.1.2 Look for ‘similar’ websites

Typosquatting is when you create a site whose url is *almost* the same as a real or well-known one, often using combinations of letters (e.g. ‘nn’ instead of ‘m’) or urls (e.g. .gov.us) to fool people on a casual glance.

Useful python libraries for generation typosquats include dnstwist

- Near-duplicates https://github.com/justinnbt/SnaPy
- https://github.com/topics/typosquatting

Feed of domains that were created each day: https://whoisds.com/newly-registered-domains

- Idea: we could search this feed each day for domain names matching the things of interest to us, e.g. MMS

8.1.3 Look for new sites

Github code check_new_registrations.ipynb [Comment24] searches for strings of interest in newly-registered domains (from https://whoisds.com/newly-registered-domains). But newly registered alone isn't really an indication of anything; domains that are newly registered and active all within 24hrs, are worth watching, as is any recently active and questionable domain. We have e.g. the Zetalytics API for searching through those.

8.1.4 Social media references to the site

Crowdtangle chrome extension will give you a list of references to a site you’re looking at, on Facebook, Twitter, Instagram and Reddit https://apps.crowdtangle.com/chrome-extension

8.2 Handling Tweets

8.2.1 Chasing a hashtag

“what do we consider worthy of collecting from twitter?” - FrankC

Good question. The TL;DR is that the reason we use the code that we do (andypatel_gettwitter.py from CSC tracking repo) is because we’re looking for the objects that dominate and are related to the hashtag:

- we want to know which users are promoting it
- Which other hashtags are used heavily with it
- Which users on the hashtag are in suspicious configurations - e.g. one user linked out to lots of other people who aren’t connected to each other (that’s someone either pushing or pulling, depending on the direction of the links), or groups of users connected heavily to each other but not to anyone else on that hashtag (typical configuration for a botnet)

- we want to know which URLs are associated with the hashtag - if this is being used to make money, that money has to come from somewhere, and that's usually either online advertising, merchandise or paid services: either way, each of those is going to have a web address associated with it, and any grifter worth their salt is going to be pushing that address heavily
- We also collect images - that gives a good idea of what the themes are, because most good disinformation merchants know that images are more often exchanged than text. That's why you see all those posters with text on

The finding the configurations part - we use Gephi [comment25] to look at the network; botnets and distributors stand out like little flowers in a Gephi network. But we could use networkx to do the same thing. There are also a set of tools in OSOME that will help you examine relationships quickly. [comment26]

Raw data is useful too - it's where we start. But really, in social engineering, it's the relationships that count.

8.2.2 Chasing botnets

I use bot sentinel [comment27] and tools like it - ones like Hamilton68 monitor accounts from nation state actors (Russia, China etc - think embassy twitter feeds, RussiaToday etc), ones like Botsentinel monitor accounts active in earlier campaigns that might or might not be bots. The most valuable thing they give you is trends: what the recent chatter online is.

Bot detection is an art now. Once upon a time, it was as easy as "there are 100 accounts posting all the time, and they're all posting the same text", and finding them was basically "look for the Qanon hashtags". Now it's more subtle. There are some rules of thumb, like being suspicious of anything tweeting more than 100 times a day, but there's more to it, and a bunch of tools to help. [comment28]

8.3 Chasing an image

There are a few things you're going to want to do with an image:

- Extract the text from it
- See where else it exists online
- Check to see if it's been altered / is fake

Extracting text: You can usually extract text from images using OCR (optical character recognition). There are libraries like Tesseract [comment29] that can be called from Python (as e.g. pytesseract), but they have mixed results. A more reliable way to do this is to use the OCR built into search engines to pull the text from each image: yandex.com appears to be best at this (although always check because OCR still doesn't produce perfect results) but is Russian: if that's an issue for you, bing.com image search does this too.

Seeing where else an image is online:

- Mostly you'll be doing this by hand for new images, but a good first check is to see if an image (e.g. a photo) has been reused from an earlier event. Reverse image search from yandex.com and bing.com works well - tineye.com will call all the big image search engines for you (and you can laugh at some of the things they return…).

Checking for alterations: Bellingcat are the masters of online image forensics, and have a good guide to this (Bellingcat guide). Look at tools like FotoForensics.

8.4 Handling Video and Audio

8.4.1 Checking video

https://twitter.com/InVID_EU

8.4.2 Save an audio file from Facebook Messenger

The workaround is:

- Using Chrome browser (but NOT on mobile)
- Access facebook via m.facebook.com
- Then click on the messenger icon
- Go to the chat that has the audio
- Right mouseclick on the (...) at the end of the message and you'll have the option to "Save Audio As"

8.5 Searching through Facebook Groups

A lot of Covid19 disinformation is happening and/or moving at some point through facebook groups. We've been tracking some of these by hands whilst working out how to automate creating watchlists of groups, pages, accounts to check for new disinformation incidents forming before they hit the mainstream press.

Some academic references on this, focussed on antivax (one of the best-known and well-studied modern conspiracy theories)

- The online competition between pro- and anti-vaccination views
    - Hidden resilience and adaptive dynamics of the global online hate ecology | Request PDF

    - https://science.sciencemag.org/content/352/6292/1459 with https://science.sciencemag.org/content/sci/suppl/2016/06/15/352.6292.1459.DC1/Johnson-SM.pdf
-

9 Making analysis outputs usable

Data science, data analysis, starts and ends with human beings. We can do beautiful analysis, but if we don't make it accessible to the people who need to take action from it, then we haven't done our job.

Let's talk about outputs. The ways we present the data we produce, and how we do that, including the forms/ formats some of the people we interact with are used to, what good visualisations in this space look like (and how to create them), and how to get those outputs to the right people.

10 Taking Action

The point of real-time disinformation tracking is to be able to do something about it. Our basic actions include:

- Direct action.
- Asking someone directly connected to us to take action
- Reporting to someone not directly connected to us, so they can investigate and decide whether to take action.

Direct action: there are many small things that a team could do to disrupt a disinformation incident. These include:

- Flooding a disinformation hashtag or group with alternative information (be careful with this because if the original intent was confusion, you might be adding to it) etc

Asking someone connected to us to take action

- Reporting a suspicious domain to registrars. If we do this, it's on us to gather information to help them - e.g. screenshots of selling bleach 'cures' etc etc

Reporting to someone not directly connected

- This is most likely with the large social media platforms. We're going to find bots and botnets; we won't be able to remove them ourselves, we will be able to report them to platforms. It'll help if we have that reporting mechanism set up ahead of time.

10.1 Reporting

10.1.1 Reporting inside the League

If you know which organisation you need, use the /list_orgs and /list_contacts [org] slack command to find the person you need. More generally, look at the channels guide in the League handbook to see the right channel to report an incident or component to.

10.1.2 Reporting to law enforcement from the League

You can open an LE escalation ticket using the /lenew command

10.1.3 Reporting to platforms

Reporting to social media

- Reddit: https://www.reddit.com/r/redditsecurity/
- Twitter: report-twitter-impersonation and twitter-rules
- Facebook: How to Report Things on Facebook
- Linkedin: Reporting Inaccurate Information on Another Member's Profile

- Instagram: https://help.instagram.com/1735798276553028
- YouTube: https://support.google.com/youtube/answer/2802027
- Google: is going to take some digging Avoid and report Google scams - Google Help

Pinterest

- Fast: https://help.pinterest.com/en/article/report-something-on-pinterest
- Slower: report on https://help.pinterest.com/en/contact?page=about_you_page - you'll need a Pinterest account to do this from.
    - Choice is porn, violence, hate speech, self harm, harassment/ exposed private information, spam; currently going with either hate speech, violence or harassment as appropriate.
    - Has an image filesize limit of 2MB
- community guidelines are https://policy.pinterest.com/en-gb/community-guidelines:
- "Hateful activities. Pinterest isn't a place for hateful content or the people and groups that promote hateful activities. We limit the distribution of or remove such content and accounts, including:
    - Slurs or negative stereotypes, caricatures and generalisations
    - Support for hate groups and people promoting hateful activities, prejudice and conspiracy theories
    - Condoning or trivialising violence because of a victim's membership in a vulnerable or protected group
    - Support for white supremacy, limiting women's rights and other discriminatory ideas
    - Hate-based conspiracy theories and misinformation, such as Holocaust denial
    - Denial of an individual's gender identity or sexual orientation, and support for conversion therapy and related programmes
    - Attacks on individuals including public figures based on their membership in a vulnerable or protected group
    - Mocking or attacking the beliefs, sacred symbols, movements or institutions of the protected or vulnerable groups identified below
- Protected and vulnerable groups include: people grouped together based on their actual or perceived race, colour, caste, ethnicity, immigration status, national origin, religion or faith, sex or gender identity, sexual orientation, disability, or medical condition. It also includes people who are grouped together based on lower socio-economic status, age, weight or size, pregnancy or ex-military status.
- Misinformation. Pinterest isn't a place for misinformation, disinformation or mal-information. We remove or limit distribution of false or misleading content that may harm Pinners' or the public's well-being, safety or trust, including:
    - Medically unsupported health claims that risk public health and safety, including the promotion of false cures, anti-vaccination advice, or misinformation about public health or safety emergencies
    - False or misleading content about individuals or protected groups that promotes fear, hate or prejudice

    - False or misleading content that encourages turning individuals, groups of people, places or organisations into targets of harassment or physical violence
    - Conspiracy theories
    - False or misleading content that impedes an election's integrity or an individual's or group's civic participation, including registering to vote, voting and being counted in a census
    - Content that originates from disinformation campaigns
    - Factual information that's published or deliberately modified to erode trust or inflict harm, such as changing or omitting context, date or time
    - Fabricated or meaningfully manipulated visual or audio content that erodes trust or causes harm
- Harassment and criticism. Pinterest isn't a place to insult, hurt or antagonise individuals or groups of people. There are good reasons to express criticism, but we may limit the distribution of or remove insulting content to keep Pinterest a positive, inspiring place; this includes:
    - Manipulated images intended to degrade or shame
    - Shaming people for their bodies or assumed sexual or romantic history
    - Sexual remarks about people's bodies and solicitations or offers of sexual acts
    - Criticism involving name-calling, profanity and other insulting language or imagery
    - Mocking someone for experiencing sadness, grief, loss or outrage
- We've also put together some resources (opens in a new window) for you to protect yourself."

10.2 Direct Action

11 Tools

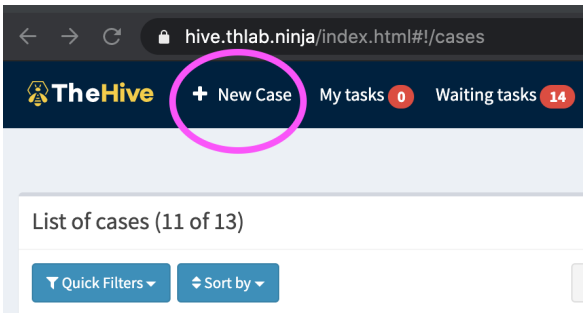
11.1 HIVE

We use Hive to manage our list of incidents, and links from them to the other objects and data connected to incident responses.

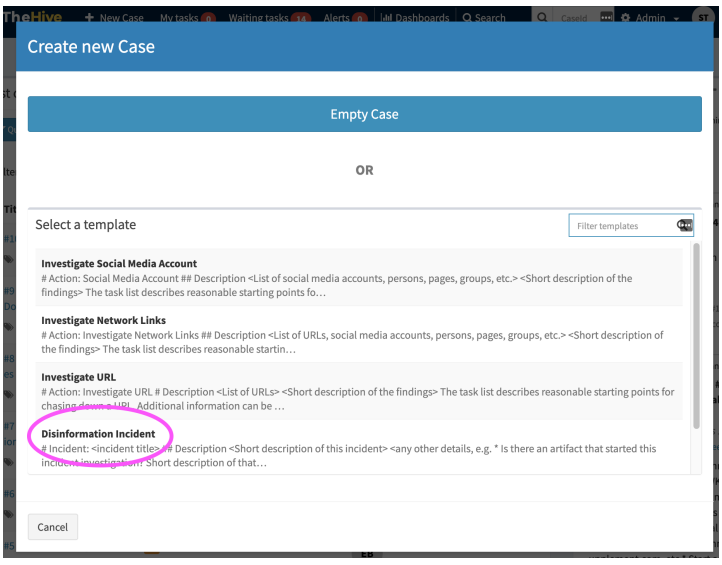
11.1.1 Adding an Incident to HIVE by hand

First, check that this incident isn't already in Hive. Check https://hive.thlab.ninja/index.html#!/cases and search for the incident name. All incidents will have the tag "disinformation" and word "Incident" in the title, which should help with searching. If you can't find this incident:

1. Click on "New Case" (the top bar, to the left)



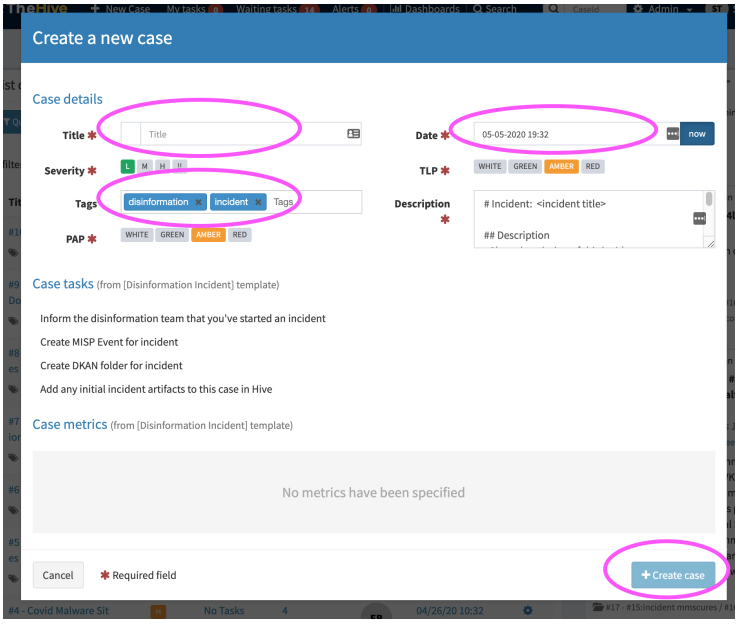
2. Select template "Disinformation Incident"



3. Fill out:

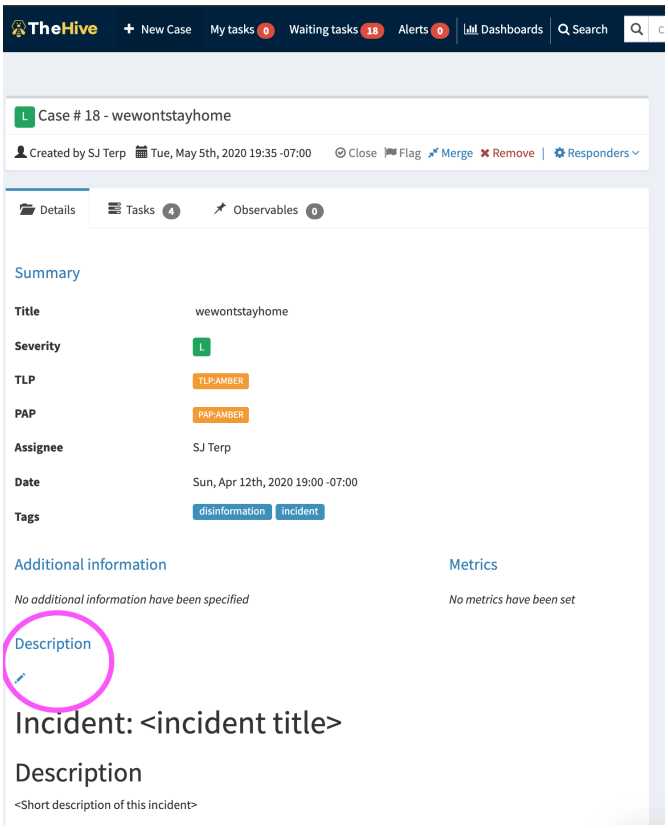
- **Title**: the name of the incident. Short, descriptive (i.e. the name you put in the incidents spreadsheet)
- **Tags**: any new tags you think this incident needs, e.g. Covid19, 5g, mms etc
- **Date**: the date we first found this incident.

4. Click "**create case**"

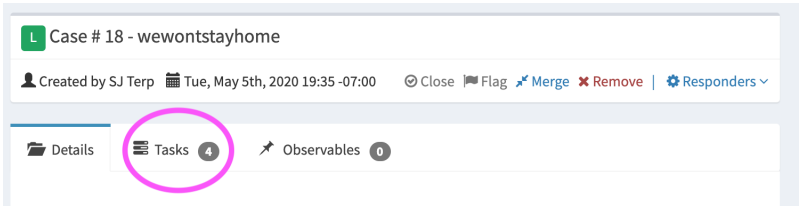


5. Now you have an incident. You still have some cleaning to do.
- Click the pencil symbol below "Description"
- Replace everything in <> with the relevant item, e.g. the incident title, description, any details you have about it

6. Also please please put the Hive number into the incidents spreadsheet, in the "Hive ID" column for this incident.



7. Not quite done - one of the things that Hive does is give you a list of tasks to do. At the top of the incident case, you'll see a "Tasks" tab. Click on it. Do the tasks (or leave them for the next person to pick up).



Congratulations - you just started an incident

11.1.2 (Adding an object workflow to a Hive Incident - don't use this yet)

Adding a new workflow to a case:
1. Assume the current Case ID is (A).
2. Create a new Case (B) selecting the workflow Case Template you wish to add to Case (A).
3. Open Case (A) and click "merge".
4. Select "By Number" and add Case ID (B).

11.2 MISP

Our main MISP instance is **https://covid-19.iglocska.eu** - we share this with the whole of the CTI League.

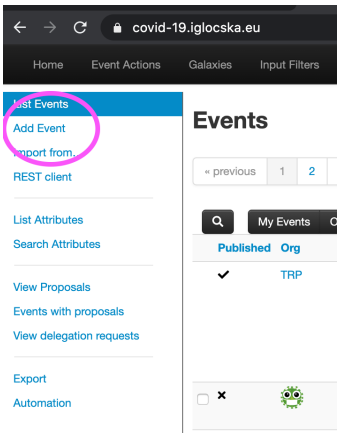
https://bbb.secin.lu/b/ale-q6v-ecn <-- Recorded MISP Training for COVID courtesy the CIRCL folks

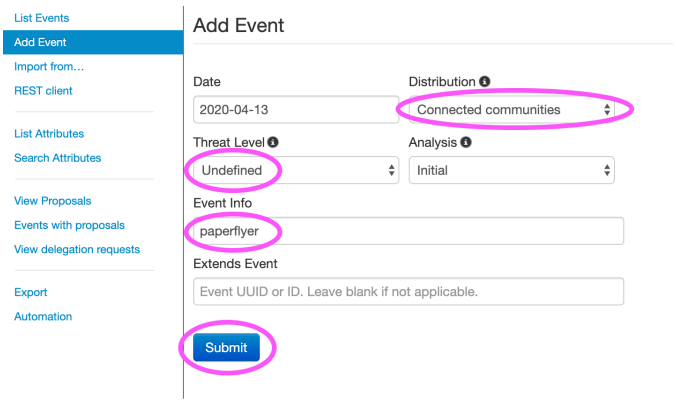
11.2.1 Adding an incident to MISP by hand [Comment30]

- Go to MISP **https://covid-19.iglocska.eu**
    - If you don't have an account yet, sign up at https://covid-19.iglocska.eu and register with the organisation "covid-19".
- Add a new incident to the MISP
    - Event Actions -> Add Event
    - Set distribution = "connected communities"
    - Set threat level = undefined
    - Set event info = name of incident
    - Hit "submit"
- Next page: add details to the new incident
    - Set tags = "current-event:pandemic="covid-19"" (in global -> pandemic)
    - Set tags = "pandemic:covid-19="disinformation"" (in global -> pandemic)
    - Click inside Galaxies -> global -> misinfosec -> Misinformation pattern
    - Click on techniques you're seeing in this incident (you can add more later)
    - Scroll up, and click on "submit"
- You now see the incident listed.

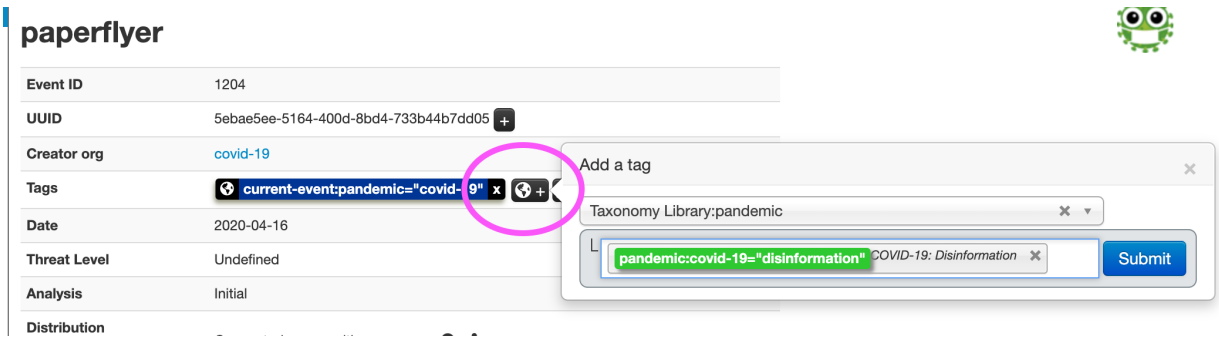
With pictures:

- Add an event:

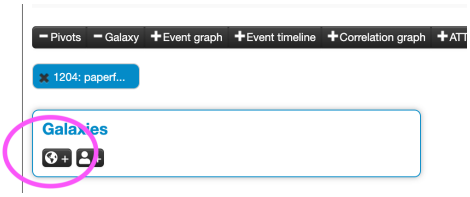


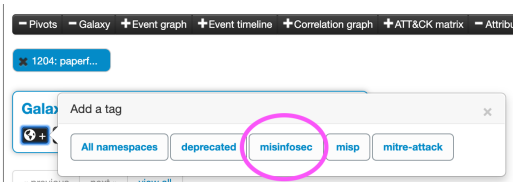


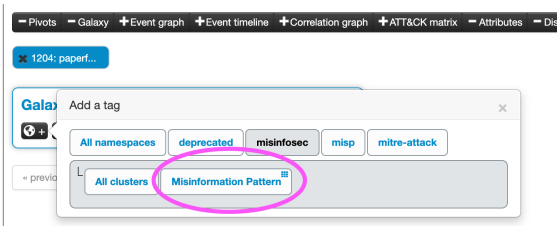
* Add tags to the event:



- Add AMITT techniques to the event:







11.2.2 Adding an object (tweet etc) to MISP by hand

- Go to MISP **https://covid-19.iglocska.eu**
    - Click on the incident ID in the list of events.
- Click on “Add Object” in the left-side column
    - Misc -> microblog for twitter or Facebook posts
    - Fill out the details
    - Click submit
    - Repeat for more objects
- Now you can start playing with the grey bar at the bottom of the event description, and toggle things like the timeline on and off.

Object types we’re most likely to need are:

Object	**Misp**	**Hive equivalent**
Facebook group	misc:facebook-group	url
Facebook page	misc:facebook-page	url
Facebook account	misc:facebook-account	url
Facebook post	misc:facebook-post	url
Twitter account	misc:twitter-account [Comment31]	url
Twitter list	misc:twitter-list [Comment32]	url
Twitter post	misc:twitter-post [comment33] (was misc:microblog)	url
Blogsite	network:url	url
Blog account	misc:user-account	url
Blogpost	misc:blog	url
Reddit group (subreddit)	misc:reddit-subreddit	url
Reddit account	misc:reddit-account	url
Reddit post	misc:reddit-post	url
Reddit post comment	misc:reddit-comment	url
YouTube Channel	misc:youtube-channel	url
YouTube Video	misc:youtube-video	url

YouTube Playlist	misc:youtube-playlist	url
YouTube Comment	misc:youtube-comment	url
Website address	network:url	url
Hashtag	ADD NEW	hashtag
Instant message	misc:instant-message	
Instant message group	misc:instant-message-group	
Narrative	misc:narrative	
Image	file:image	
Meme	file:meme-image	
Individual	misc:person	
Event [comment34][comment35] (e.g. protest)	misc:scheduled-event	
Location	misc:geolocation	

Other objects we might need include:

Object	**Misp**	**Hive equivalent**
	misc:course-of-action	
	network:email	
	file:forged-document	
	file:leaked-document	
	misc:legal-entity	
	misc:news-agency	
	misc:organization	
	misc:scheduled-event [comment36]	
	misc:short-message-service	

	[Comment37][Comment38]	
	network:shortened-link	
	misc:user-account [comment39]	

11.2.3 Adding an object to MISP via Slack bot

- Slack bots can quickly create and append an object to an event.
- Each bot attempts to modify the MISP event directly. If it lacks permission it will instead create a MISP event extension. Click the icon shown below to switch to extended mode to see the extended event objects appended into the main event.

- 

11.2.3.1 Twitter Posts

There's a Slackbot in #4-disinformation that can upload a Twitter post to a MISP event. The bot works like this /misp_twitter $MISP_event_id $post_id
It accepts either a Twitter Status ID or a Twitter post URL as arguments for $post_id

- In the #disinformation channel use the following command to add a Twitter post to the CTI League MISP
  - /misp_twitter <misp event id> <twitter post URL or twitter post ID>
  - Example: /misp_twitter 34 https://twitter.com/NASA/status/1259960728951365633?s=20

11.2.3.2 BuiltWith Tags

- In the #disinformation channel use the following command to add a Twitter post to the CTI League MISP
  - /misp_builtwith <misp event id> <url or domain name>
  - Example: /misp_builtwith 34 newyorkcityguns.com

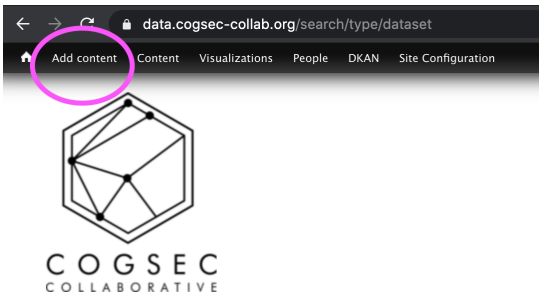
11.3 DKAN

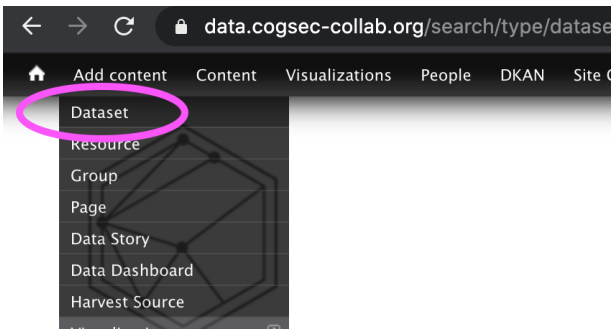
DKAN is a data warehouse tool - it's where we store large datasets and their descriptions,for analysts to use.

11.3.1 Creating a dataset in DKAN

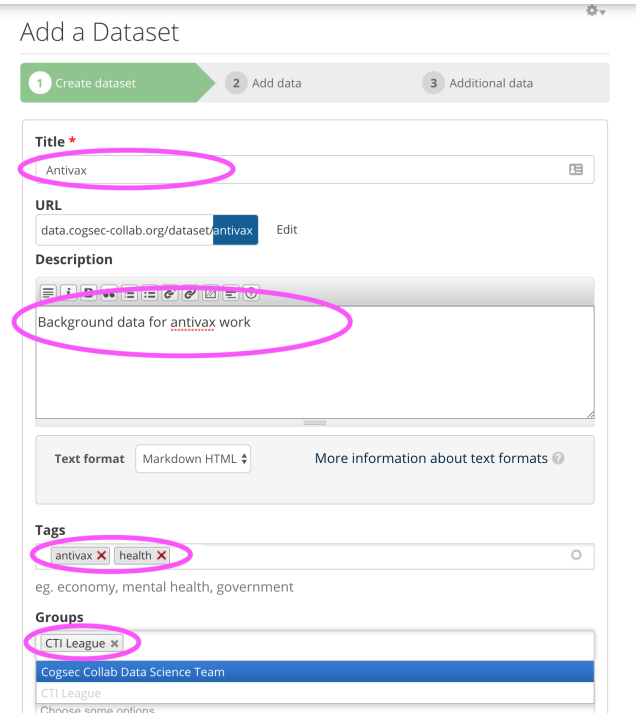
To create a **dataset** in DKAN, e.g. as somewhere to put large datasets associated with an incident, do this:

1. Log into DKAN. On the top left of the screen, there's an "add content" button. Click it, then click "dataset".

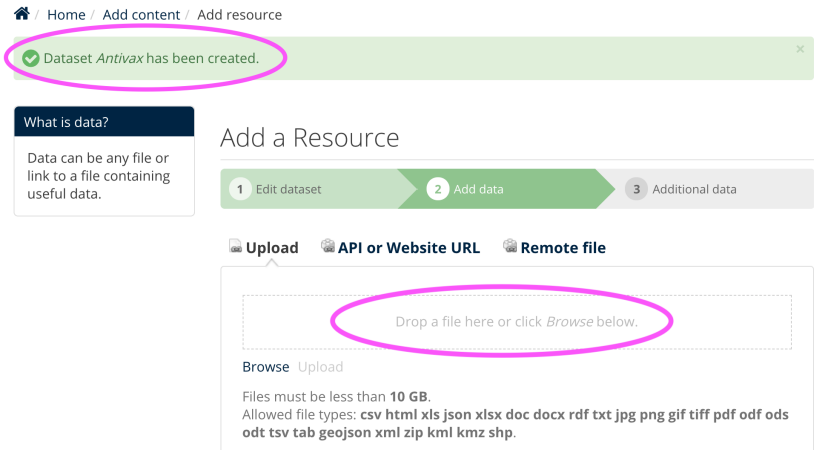




2. Now you'll see the "Add a dataset" screen. Give this dataset a name, add a description of it, some tags (to make it easy to find later), assign it to some groups (so people can see it), and give the dataset a license so people can use it (most of my data is CC0 licensed - your license choices might differ).



3. Yay. Now you've got a dataset. You can either leave it like that ready for other people to add data to it, or start adding data files to it yourself, using the handy "drop a file here" area below.



11.4 Gephi

11.4.1 Viewing networks with Gephi

This is a manual process with instructions created from Andy Patel's video at https://www.youtube.com/watch?time_continue=17&v=AqlT0khVuZA

- Get Gephi from https://gephi.org/users/download/ - install it.
- Start Gephi.
- Click on top menu>file>“import spreadsheet”. Grab User_user_graph.csv - use all defaults
- Top menu: Go to data laboratory, “copy data to another column”, click ‘id’, click okay.
- Go to overview. RHS: Run modularity algorithm, using defaults
- RHS: Run average weighted degree algorithm
- LHS: Click color icon, then partition, modularity class. Open palette, generate, unclick “limit number of colors”, preset=intense, generate, okay
- LHS: Select “tt”, ranking, weighted degree, set minsize=0.2, choose 3rd spline, apply
- LHS: Layout: OpenOrd, run. Then forceatlas2, run. Try stronger gravity, and scaling=200
- Top menu: Preview - select “black background”, click “refresh”. Click “Reset zoom”

Gephi has an API - these tasks could be automated.

11.5 Slack bots

We use slack bots to push artefacts to MISP.

we can now add the following object to a MISP event using the following slash commands
/misp_reddit_account
/misp_reddit_post
/misp_reddit_comment
/misp_reddit_subreddit

If we want new ones - we can build them, and [Name Redacted] wrote a handy how-to guide: https://vvx7.io/posts/2020/05/misp-slack-bot/

If we want new MISP object types, here’s how to do that too: [comment40]

1. Create the new object folder
    a. Git clone https://github.com/MISP/misp-objects
    b. Go into repo folder objects. It contains a subfolder for every misp object type
    c. Copy one of the existing object folders; rename the copy to the new object you want
    d. Go into the new object’s folder. You’ll find one file in here: definition.json. Open it for editing
2. Set basic data
    a. Get a new UUID from https://www.uuidgenerator.net/ - replace “uuid” in definition.json with this new one
    b. Set “version” to 1
    c. Set “name” to the same as the new folder name (nb use “-” not “_”)
    d. Set “description” to something descriptive
    e. “Meta-category” is usually “misc”

3. Set attributes. Go through attributes. For each one, set:
    a. “Description”: something descriptive
    b. “Misp-attribute”: see https://www.circl.lu/doc/misp/categories-and-types/. You’ll probably use “text” a lot. The difference between url and link? url isn’t trusted; link is trusted (this signals whether something is safe to click on).
    c. “Ui-priority”: just leave this as default (1 is always okay)
4. These attributes aren’t mandatory, but are useful
    a. “Multiple”: set this to “true” if you allow multiple of this attribute (e.g. hashtags)
    b. "disable_correlation": true, - stops MISP trying to correlate this attribute - set this on things like language to stop MISP from wasting time
    c. "to_ids" - makes exportable via api - set to false as needed (most attributes don’t need it)
5. Set the list of attributes that an object must have one of to exist
    a. List these in “requiredOneOf”
6. Check the new object is valid
    a. Run validate_all.sh
    b. Run jq_all_the_things.sh
7. Push your change back to the MISP objects repo (or to [Name Redacted] for sanity-checking)

11.6 Python scripts [Comment41]

We use python a lot (just look at the github repo…). Here are some useful resources:

- Learn python the hard way
- ACTION: [Name Redacted] add notes on python and data science - [Name Redacted] - level friendly

11.7 Other Tools

We’ve mentioned a bunch of tools above.

Some basic tools:

- Most data scientists use Python and Jupyter notebooks [comment42]. You’ll see a lot of these - the basic Anaconda install comes with most of the things we use https://www.anaconda.com/distribution/
- Data gathering:
    - Reaper https://github.com/ScriptSmith/reaper https://github.com/ScriptSmith/socialreaper https://reaper.social/ - scrapes Facebook, Twitter, Reddit, Youtube, Pinterest, Tumblr APIs
- Network analysis and visualisation: there are many tools for this.
    - Gephi is a good standalone tool https://gephi.org/users/install/
    - Networkx is a useful python library
- URL analysis
    - Builtwith.com

- Image analysis
    - Reverse image search: tineye.com, Bellingcat guide
    - Image search: bing.com, yandex.com
    - Image text extraction: bing.com, yandex.com
- Data storage / Threat Intelligence tools
    - DKAN https://getdkan.org/
    - MISP https://www.misp-project.org/

Disinformation-specific tools:

- Indiana University has a set of tools at https://osome.iuni.iu.edu/tools/
    - Botometer: check bot score for a twitter account and friends https://botometer.iuni.iu.edu/#!/
    - Hoaxy: check rumour spread (uses Gephi) https://botometer.iuni.iu.edu/#!/
    - Botslayer https://osome.iuni.iu.edu/tools/botslayer/
- Bellingcat made a list of useful tools
    - Bellingcat's really big tools list - worth reading if you need a specific OSINT tool

12 References

12.1 CTI Disinformation Reading Group

We have a reading group! We meet Fridays at 4pm PST/ 7pm EST. [Name Redacted] is running the group.
https://us02web.zoom.us/j/85454323080?pwd=bjNnVE1RY3pmSVhFaWd3UXR1YUVHUT09

12.1.1 Reading Schedule

- Friday May 8th 2020: ARTICLE Facebook shut down commercial disinformation network based in Myanmar and Vietnam
- Friday May 15th 2020: **NO READING GROUP MEETING**
- Friday May 22nd 2020: BOOK Thomas Rid's "Active Measures" Chapter 1 & 2
- Friday May 29th 2020: ARTICLE (Bellingcat) Uncovering A Pro-Chinese Government Information Operation On Twitter and Facebook: Analysis Of The #MilesGuo Bot Network
- Friday June 5th 2020: ARTICLE Unpacking China's Viral Propaganda War
-

12.1.2 Meeting Notes

- Friday May 8th 2020 Notes

12.2 Bedtime Readings

12.2.1 Books

If you really want to get into how we got here, the history of information operations, what disinformation and propaganda are etc, these books were recommended by the team:

- [Name Redacted]'s 2018 book stack - dated, but some good classics in here
- Thomas Rid's "Active Measures"
- PW Singer and Emerson Brooking's "Like War"
- Zeynap Tufeki's "Twitter and Tear Gas" (free version)
- Verification handbook: handbook, investigative reporting
- Verification Handbook: homepage

12.2.2 Articles

- Unpacking China's Viral Propaganda War
- Prevalence of Low-Credibility Information on Twitter During the COVID-19 Outbreak (5 pages)
- Media Manipulation and Disinformation Online (106 pages)
- Facebook's Coordinated Inauthentic Behavior - An OSINT Analysis
- Naval Post Graduate - Disinformation (many)
- Hate multiverse spreads malicious COVID-19 content online beyond individual platform control (9 pages)
- From Russia with Blogs (26 pages)
- The COVID-19 Social Media Infodemic (18 pages)
- We've Just Seen the First Use of Deepfakes in an Indian Election Campaign
- Facebook shut down commercial disinformation network based in Myanmar and Vietnam
- Facebook April 2020 Coordinated Inauthentic Behavior Report (26 pages)
    - Iran's Broadcaster: Inauthentic Behavior (46 pages)
    - Facebook's VDARE Takedown (18 pages)
    - Facebook Downs Inauthentic Cluster Inspired by QAnon (19 pages)
- (Bellingcat) Uncovering A Pro-Chinese Government Information Operation On Twitter and Facebook: Analysis Of The #MilesGuo Bot Network
- Unmaking Democracy: How Corporate Influence Is Eroding Democratic Governance (Harvard International Review) - 4 May 2020 (6min read)
- Conspiracy Theory Handbook (12 Pages)
    - Google Drive Location
- What if we've all been primed? (6 pages)
- (Bellingcat) Investigate TikTok Like a Pro (15min read)

12.2.3 Podcasts and videos

- Motherboard's Cyber Podcast Episode with Thomas Rid about Active Measures and implications for modern disinformation
- Lawfare's Arbiters of Truth podcast series about disinformation
    - Especially this episode with Camille Francoise specifically about COVID-19 disinfo and the ABCs of Disinfo
- vOPCDE #2 - Discussion: Disinformation about Disinformation ([Name Redacted], [Name Redacted], [Name Redacted])
-

12.2.4 People to Follow

Disinformation data science:

- Conspirador Norteno and Dr ZQ: @conspirator0 @ZellaQuixote
    - always a great example on bot tracking
    - went looking at reopen etc https://twitter.com/conspirator0/status/1252374902121721859?s=19
    - Tools: https://makeadverbsgreatagain.org/allegedly and python/jupyter with libraries pandas, tweepy, bokeh, cytoscape
    - Looking at a botnet: https://twitter.com/conspirator0/status/1265829648056954881?s=20
- Andy Patel: @r0zetta
    - Infosec and misinformation data scientist
    - Tools: e.g. using TFIDF plus Louvain clustering to analyse twitter https://twitter.com/r0zetta/status/1230786764413030400 https://twitter-clustering.web.app/ (https://github.com/r0zetta/meta_embedding_clustering)
    - https://blog.f-secure.com/author/andrew-patelf-secure-com/page/2/
- Elliot Alderson: @fs0c131y (fs0c131y.com)
    - Infosec and misinformation data scientist

Disinformation tracking:

- Erin Gallagher: @3r1nG
- @josh_emerson
- Kate Starbird: @katestarbird

12.2.5 Examples of disinformation tracking

- "Distinguished Impersonator" Information Operation That Previously Impersonated U.S. Politicians and Journalists on Social Media Leverages Fabricated U.S. Liberal Personas to Promote Iranian Interests
- From Russia With Blogs
- Facebook shut down commercial disinformation network based in Myanmar and Vietnam

- Facebook's Coordinated Inauthentic Behavior - An OSINT Analysis

Disinfo data science (short investigations)
- https://onezero.medium.com/facebook-groups-and-youtube-enabled-viral-spread-of-plandemic-misinformation-f1a279335e8c

Images and disinformation
- Deepfakes by BJP in Indian Delhi Election Campaign

Disinformation Counters
- Training end-users about disinformation
    - https://getbadnews.com/#intro - game to train people on how disinformation works
    - CrashCourse media literacy videos

12.3 Covid19 disinformation references

Disinformation
- https://tomnikkola.com/prime/
- https://www.russiamatters.org/analysis/dark-arts-disinformation-through-historical-lens
- https://www.brookings.edu/blog/order-from-chaos/2020/05/19/the-kremlins-disinformation-playbook-goes-to-beijing/
- https://www.snopes.com/news/2020/05/20/michigan-lockdown-protesters/
- https://eng.lsm.lv/article/society/defense/nato-stratcomcoe-considers-disinformation-in-asia.a360361/
- https://www.atlanticcouncil.org/blogs/new-atlanticist/activists-fight-covid-19-disinformation-in-the-caucasus/
- https://www.aljazeera.com/indepth/features/anatomy-disinformation-campaign-coup-200518142503624.html
- https://hub.jhu.edu/2020/05/08/thomas-rid-disinformation-in-covid-19-pandemic/
- https://warontherocks.com/2020/05/disarming-disinformation/
- https://www.expressandstar.com/news/uk-news/2020/05/20/tech-giants-recalled-by-mps-over-lack-of-adequate-answers-on-disinformation/
- https://www.businessinsider.com/coronavirus-trump-us-disinformation-foreign-interference-2020-4
- https://www.foxnews.com/politics/house-democrats-coronavirus-bill-1-million-study-disinformation
- https://www.thetelegraph.com/opinion/article/Editorial-Disinformation-15279496.php
- https://www.atlanticcouncil.org/event/belarus-moldova-and-ukraine-covid-19-disinformation-in-eurasia/
- https://www.theverge.com/2020/5/12/21255823/alex-stamos-interview-election-2020-security-tik-tok-zoom-vergecast

- https://www.businessinsider.com/coronavirus-trump-us-disinformation-foreign-interference-2020-4
- https://www.oregonlive.com/nation/2020/05/eu-tackles-coronavirus-disinformation-seeks-regulatory-framework-for-facebook-other-social-media-companies.html
- https://www.ft.com/content/e479043f-a197-41f1-86c8-ff6c922e3492
- https://www.winchesterstar.com/coronavirus/wexton-seeks-study-of-covid-19-disinformation-misinformation/article_29119442-31ab-5e76-b362-9e65bcfe6141.html
- https://www.bellingcat.com/news/2020/05/05/uncovering-a-pro-chinese-government-information-operation-on-twitter-and-facebook-analysis-of-the-milesguo-bot-network/
  Also briefly about Covid19 disinformation network

Covid19 Narratives

- https://allianceforscience.cornell.edu/blog/2020/04/covid-top-10-current-conspiracy-theories/

December 2023: comments on document (moved here to protect PII)

- [Comment01] What tool set would need to be developed/deployed to facilitate this?
- [Comment02] What tool set would need to be developed/deployed to facilitate this?
- [Comment03] This section is good, but we need to make the language more uniform
- [Comment04] his is super important and has important implications for our slack channel structure and our management of our channels. How can we most effectively make these norms visible to folks coming to join our Slack team?
- [Comment05] Tidy up the language here to match the style of the rest of the document
- [Comment06] Turn this into prose
- [Comment07] I think this should be like a board to determine, i am thinking maybe one of the leads posting it to #disinformation-triage. It should include the link to the thread (with the most evidence) and start it with a key selector (i.e. VOTE or START??)
- [Comment08] Possibly add where we can observe current or NEW alerts for each stream
- [Comment09] And maybe if they are accessible to the CTI League members, admins, managers, or reliant on external
- [Comment10] We could also maybe create like a checklist, that everyday by XX:XX every day, each source is checked and etc
- [Comment11] Link?
- [Comment12] Link?
- [Comment13] ?
- [Comment14] Isn’t this what triage is for?
- [Comment15] The level of detail here is awesome. Would be next level to actually have an exemplar built out that we can refer back to. Don't know if we have an incident we've

taken all the way through the process that we can do this for or if that will be forthcoming.

- [Comment16] Add link to Team Log folder
- [Comment17] We should move to exclusive use of TheHive for this so that we don't have to deal with the scaling issues after it's too late
- [Comment18] Link to folder or shift to Hive as above
- [Comment19] Have we gotten to this stage yet with an incident in the CTI League?
- [Comment20] Which feeds?
- [Comment21] This is the third time we've redefined narratives.
- [Comment22] Wired resolve issue, resolved URL on one but not the other
- [Comment23] Another redefinition
- [Comment24] Add link to this code
- [Comment25] Link to Gephi
- [Comment26] Add links here
- [Comment27] Add link here
- [Comment28] Link to these tools
- [Comment29] Add link
- [Comment30] Flesh this out in similar depth to the TheHive section
- [Comment31] Created, not in MISP yet
- [Comment32] created, not in MISP yet
- [Comment33] created, not in misp yet
- [Comment34] we have a lot of disinfo around events, e.g. protests - how do we represent?
- [Comment35] One way we can do this is with tags/taxonomies. A "protest" or "earthquake" tag make sense and they're nice and generic.
- [Comment36] See note above about events
- [Comment37] Is this SMS as a service, or SMS as a message? We'd definitely need an SMS message object
- [Comment38] This is for a message.
- [Comment39] Can we use this for social media accounts, or is this network / pc accounts only?
- [Comment40] Flag these in kanban as opportunities to build automation tooling
- [Comment41] Add to kanban as an area for volunteer involvement
- [Comment42] Link to jupyter notebooks primer