Hybrid CoE Research Report 7

# Foreign information manipulation and interference defence standards: Test for rapid adoption of the common language and framework 'DISARM'

Hybrid CoE

Hadley Newman – November 2022

**Hybrid CoE Research Reports** are thorough, in-depth studies providing a deep understanding of hybrid threats and phenomena relating to them. Research Reports build on an original idea and follow academic research report standards, presenting new research findings. They provide either policy-relevant recommendations or practical conclusions.

This report has been produced in collaboration with the **NATO Strategic Communications Centre of Excellence**.

**The European Centre of Excellence for Countering Hybrid Threats**
tel. +358 400 253800 | www.hybridcoe.fi

ISBN 978–952–7472–46–0 (web)
ISBN 978–952–7472–47–7 (print)
ISSN 2737–0860 (web)
ISSN 2814–7219 (print)

November 2022

Cover photo: Bits And Splits / Shutterstock.com

**Hybrid CoE's mission** is to strengthen its Participating States' security by providing expertise and training for countering hybrid threats, and by enhancing EU-NATO cooperation in this respect. The Centre is an autonomous hub for practitioners and experts, located in Helsinki, Finland.

The responsibility for the views expressed ultimately rests with the authors.

# Contents

# Figures

# Tables

# Glossary

| Term | Explanation |
|---|---|
| AMITT | Adversarial Misinformation and Influence Tactics and Techniques framework for describing disinformation incidents. Includes TTPs and countermeasures (now DISARM). |
| Campaigns | Campaigns are advanced persistent threats predominantly created by nation-state actors using information manipulation and interference with long-term objectives, and consisting of multiple incidents. A campaign is an overarching term for intelligence-based information about a particular kill-chain-based intrusion and comprises intrusion attempts, combined with TTPs. |
| DISARM | DISinformation Analysis & Risk Management (formerly AMITT) is the open-source, master framework for fighting disinformation through the coordination of effective action. |
| Disinformation | Verifiably false or misleading information that is created, presented and disseminated for economic and political gain or to intentionally deceive the public, and may cause public harm – with public harm comprising threats to democratic political and policy-making processes as well as public goods such as the protection of [EU] citizens' health, the environment or security.[1] |
| FIMI | Foreign information manipulation and interference. Often labelled as 'disinformation'.[2] |
| Incidents | Sets of FIMI activity with specific objectives, e.g., to change beliefs, emotions, or behaviours. Bursts of activity may be opportunistic, and created by individuals, groups, and organizations. "Instances are described using data, such as time-related information, location of effect, related indicators, leveraged TTP, suspected intent, impact assessment, response course of action requested/taken, source of the incident information, and log of actions taken."[3] |
| Infosec | Information security is a multidisciplinary area of study and professional activity that includes the practice of protecting information by mitigating information risks. "Properties such as authenticity, accountability, non-repudiation and reliability can be involved."[4] |
| JSON | Java Script Object Notation is an open data interchange format that is easy for humans to read and write in a common data format with diverse uses in electronic data interchange.[5] |

[1] European Commission, 'Tackling online disinformation: a European Approach. COM(2018) 236 final', Brussels, 2018, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0236. [All links were last accessed on 14 September 2022, unless otherwise indicated.]

[2] European External Action Service (EEAS), 'Tackling Disinformation, Foreign Information Manipulation & Interference', Stratcom Activity Report (October 2021), https://www.eeas.europa.eu/eeas/tackling-disinformation-foreign-information-manipulation-interference_en.

[3] Sean Barnum, *Standardizing cyber threat intelligence information with the structured threat information expression* (stix) (Mitre Corporation, 2012), 1–22.

[4] ISO/IEC, *ISO/IEC 27000:2009 (E), Information technology – Security techniques – Information security management systems – Overview and vocabulary* (2009).

[5] Java Script Object Notation, json.org, 2022, https://www.json.org/json-en.html.

| | |
|---|---|
| MISP | The Malware Information Sharing Platform is an open-source software solution for collecting, storing, distributing, and sharing cybersecurity indicators and threats[6] |
| MITRE ATT&CK® | A globally accessible knowledge base of adversary tactics and techniques. |
| TAXII | The Trusted Automated Exchange of Intelligence Information is the "transport mechanism for sharing cyber threat intelligence". It is a protocol that enables the sharing of cyber threat information over HTTPS.[7] |
| TTP(s) | Tactics, Techniques, and Procedures are a key concept in cybersecurity and threat intelligence. The purpose is to identify patterns of behaviour which can be used to defend against specific strategies and threat vectors used by malicious actors.[8] A tactic is the highest-level description of this behaviour and is the activity that an actor (conducting a FIMI incident) is likely to use. Techniques give a more detailed description of behaviour in the context of a tactic and are how an actor might conduct the tactic(s). Procedures are an even lower-level, highly detailed description in the context of a technique and are the detailed steps that prescribe how to perform specific tasks.[9] |
| STIX™ | Structured Threat Information eXpression enables analysts to exchange threat information.[10] |
| UNC1151 | A suspected state-sponsored cyber espionage actor and the threat categorization of a cyber-enabled influence campaign. |

[6] MISP Open Source Threat Intelligence Platform & Open Standards For Threat Information Sharing, 2022, https://www.misp-project.org/.

[7] Cyber Threat Intelligence Technical Committee, 2022, https://oasis-open.github.io/cti-documentation/.

[8] Tag Archives: TTPs, https://zvelo.com/tag/ttp/.

[9] Chris Johnson et al., 'Guide to Cyber Threat Information Sharing', National Institute of Standards and Technology, October 2016, https://doi.org/10.6028/NIST.SP.800–150.

[10] Barnum, *Standardizing cyber threat int*elligence, 1–22.

# Executive summary

This report was commissioned to examine the DISARM framework and its suitability for rapid adoption by strategic communications practitioners (representing non-specialist users), and its credibility for specialist foreign information manipulation and interference (FIMI) threat analysts.

The report contributes to an increasing body of research on FIMI defence. FIMI incidents present a growing political and security challenge and there is a need for a common defence framework. According to the European External Action Service, a whole-of-society approach is required to increase resilience and counter FIMI. Capabilities and competences that lie within governments, civil society, and private industry should be leveraged to achieve this.

As an open-source, master framework that draws on the best practices in global cybersecurity, the DISARM framework contributes to FIMI defence through the coordination of effective action. It has been used to help practitioners, from varied disciplines and sectors, to gain a shared understanding of FIMI incidents and to quickly identify the available countermeasures.

For the purposes of this report, Operation Ghostwriter (a cyber-enabled FIMI campaign) served to exemplify the use of the framework. While Operation Ghostwriter has been attributed to multiple countries, the attack targeted citizens in Lithuania, Latvia, and Poland with narratives that were critical of NATO's presence in Eastern Europe. Operation Ghostwriter exemplifies a FIMI campaign of considerable scale and potency, with prolific use of sophisticated communications tactics and techniques.

To understand the nature of the threat from FIMI, a simple metric of *observe, investigate, and identify was applied to an incident* in the Operation Ghostwriter campaign. Next, the ABC model was used to investigate: (A) What kind of actors were involved? (B) What behaviours (activities) were exhibited? and (C) What content was created and distributed?

While there is no universal and consensual framework for how to catalogue and describe FIMI activity, steps are being taken by individual bodies to codify, prioritize, and share details about information-based threats. As noted in the Joint Statement from the US-EU Summit in 2021, when discussing the intent to build a more democratic, peaceful, and secure world:

> We plan to increase cooperation and exchange information and expertise to increase resilience against and to counter foreign information manipulation and interference, all forms of coercion including economic pressure, hybrid threats, malicious cyber activities, terrorism and violent extremism, and other common security threats.

Although the aspiration is commendable, the feasibility and effectiveness of this plan could be undermined by the lack of a shared approach.

Accordingly, the compatibility of the DISARM framework with other information security tools and practices was assessed using STIX

(structured threat information expression). To develop shared data standards that would enable the collective assessment of FIMI incidents, there was also a need to test for rapid adoption and ease of use by strategic communications practitioners with no previous DISARM experience. Separately, it was important to consider the credibility of the DISARM framework as a universal approach to accurately catalogue and plan responses to disinformation threats and attacks from the perspective of FIMI threat analysts. Thus, the framework was assessed against recognized usability heuristics and credibility principles.

For this, user capability personas were employed to determine whether the DISARM framework could be quickly adopted by a strategic communications practitioner with no previous knowledge of the framework, and to establish the credibility of the DISARM framework from the perspective of a FIMI threat analyst. With this groundwork in place, the role of governments, private industry, and academia in increasing resilience against and countering FIMI was explored, and targeted recommendations by context were offered.

The DISARM framework was found to be compatible with existing tools in the information security ecosystem. By using the DISARM framework to identify the tactics and techniques identified in the Operation Ghostwriter use case, it was simple to codify the incident in STIX. Furthermore, the framework is compliant with the nine applicable heuristic tests conducted for ease-of-use and the 10 credibility principles that were considered in the assessment.

**It is recommended that the DISARM framework be used by FIMI threat analysts and strategic communications practitioners** within government, international organizations, and institutions, platforms, academia, private industry, and civil society. **Private industry should be engaged in developing application software to record, process, and visualize information activity in accordance with the DISARM framework.** Commercializing FIMI defence practices and fostering innovation would further strengthen the collective defence against FIMI. Businesses have an invaluable role to play in supporting the wider adoption of the DISARM framework by developing software applications that can make data usable and reveal its value; they would enable stakeholders to record, process, and visualize FIMI activity. As online FIMI practices are built upon contemporary communications practices, there is an opportunity for developers to integrate components of the DISARM framework into existing automation solutions. Moreover, this could accelerate fast-track adoption of the DISARM framework amongst private industry practitioners, who are already familiar with the solutions, and thus popularize FIMI defence practices.

**It is further recommended that FIMI monitoring and analysis should be established as a communications discipline** in accordance with the 'sighting', 'response', and 'analysis' process for the introduction of robust defence teams that are applicable across government as well as in private industry. **Finally, partnering with academia and professional bodies to deliver age-appropriate tutorials on a universal**

**approach** could increase the talent pool and accelerate the mobilization of defence against the growing political and security challenges associated with FIMI.

This report established that a practitioner with a basic skillset in strategic communications would find the DISARM framework intuitive to learn and easy to use. A FIMI threat analyst can feel reassured that the DISARM framework is a credible system for a universal approach to catalogue and plan disinformation threats and attacks. As a further advantage, the framework's capability development tools can facilitate large-scale capacity-building that is beneficial for stakeholders. Since the DISARM framework was built by experts across the international security community, the result is a valuable tool that is well-considered, practical, and fit for purpose. Not only is the DISARM framework a robust and market-ready solution that is responsive to immediate threats, but it can also scale as FIMI evolves and the threat increases.

The assessment revealed several significant contributing factors to ensure the DISARM framework suitability for practical application and rapid adoption. Taking a collective whole-of-society approach and immediate action is paramount to mitigate urgent, real-world problems at scale. To achieve this, practitioners within government, international organizations, and institutions, platforms, academia, private industry, and civil society must be encouraged to adopt the DISARM framework, and to support its continued development.

# 1. Introduction

This report provides an overview and application of the DISARM framework to a case of foreign information manipulation and interference (FIMI). While information warfare is not a new threat,[11] advanced technology has enabled nefarious actors to better employ information warfare as part of geopolitical strategies.

DISARM (formerly known as AMITT) is an open-source, master framework for fighting disinformation through the coordination of effective action. The DISARM framework has been developed drawing on global cybersecurity best practices. It has been used to help practitioners, from varied disciplines and sectors, to gain a shared understanding of FIMI incidents and to quickly identify the available countermeasures.[12]

This report critically examines the DISARM framework and, in doing so, evaluates whether it is suitable for rapid adoption and ease of use by practitioners in an ever-changing and emerging discipline of counter FIMI. The report includes the assessment of DISARM attributes that are critical to its credibility for FIMI threat analysts.

To this end, the report aims to:

**i.** Identify and report a use case (Operation Ghostwriter).

**ii.** Test the DISARM framework against the use case.

**iii.** Code a sample of the framework and case study using STIX to assess DISARM's workability with other infosec tools and practices.

**iv.** Identify specialists that would make up a robust defence team.

**v.** Define user capability personas – namely, a strategic communications practitioner who can identify fundamental communication techniques and yet has no previous knowledge of DISARM, and a FIMI threat analyst whose work includes the identification, cataloguing, and response to FIMI activity.

**vi.** Identify and apply a method to determine whether DISARM can be quickly adopted by a strategic communications practitioner and to establish the credibility of DISARM from the perspective of a FIMI threat analyst.

Divided into six main sections, the report begins with an overview of information manipulation, after which the DISARM framework is introduced. In the subsequent section, the Operation Ghostwriter case study is presented. This case study was used in the evaluation of DISARM and the findings are reported in the next section. The role of governments, private industry, and academia in increasing resilience against and countering FIMI is then discussed, and recommendations are offered. The report concludes with some salient observations from the evaluation.

[11] Z. Khalilzad, J. White, & A. Marshall, *Strategic Appraisal: The Changing Role of Information in Warfare* (1999), p. 180, https://www.rand.org/content/dam/rand/pubs/monograph_reports/MR1016/RAND_MR1016.pdf.

[12] See https://www.disarm.foundation/framework.

# 2. Background

FIMI as a concept has been described by the European External Action Service as a "mostly non-illegal pattern of behaviour that threatens or has the potential to negatively impact values, procedures, and political processes. Such activity is manipulative in character, conducted in an intentional and coordinated manner. Actors of such activity can be state or non-state actors, including their proxies inside and outside of their own territory".[13] The manipulative shaping of attitudes and behaviours through information warfare to weaken public trust in democratic institutions is increasingly prevalent. With more than five billion internet users worldwide,[14] the magnitude, potency, and proliferation of FIMI activity has been amplified – with ever more complex practices emerging as a mechanism of control from within the platforms and constraint from without. Similarly, the increased adoption of social media has served to exponentially increase the forms in which the power of FIMI is exerted and the intensity of that power. Information manipulation through social media has proved to be a key factor in recent information warfare[15]
and demands a collaborative global response.

No universal and consensual framework for how to catalogue and describe FIMI activity currently exists. Steps are being taken by individual bodies to codify, prioritize, and share details about information-based threats, as noted in the Joint Statement from the US-EU Summit in 2021 when discussing the intent to build a more democratic, peaceful, and secure world: "We plan to increase cooperation and exchange information and expertise to increase resilience against and to counter foreign information manipulation and interference, all forms of coercion including economic pressure, hybrid threats, malicious cyber activities, terrorism and violent extremism, and other common security threats."[16] However, their effectiveness could be undermined by the lack of a shared approach. This report identifies the DISARM framework as the basis for such an approach.

Distinct from studies on disinformation, which are often concerned with the content of activity, this report considers the broader concept of FIMI, which is more concerned with the "behaviour of an actor" that is described through tactics, techniques, and procedures (TTPs).[17] A tactic is the highest-level description of this behaviour and is the activity that an actor (conducting a FIMI incident) is likely to use. Techniques give a more detailed description of behaviour in the context of a tactic and how an actor might conduct the tactic(s). "Procedures are an even lower-level, highly detailed description in the context of a technique"[18]

[13] See https://www.eeas.europa.eu/eeas/tackling-disinformation-foreign-information-manipulation-interference_en.

[14] Statista, 'Number of internet and social media users worldwide', 2022, https://www.statista.com/statistics/617136/digital-population-worldwide/.

[15] Hadley Newman, 'Understanding The Differences Between Disinformation, Misinformation, Malinformation and Information – Presenting The DMMI Matrix', evidence to UK Government Joint Committee for the Draft Online Safety Bill, 2020, https://publications.parliament.uk/pa/jt5802/jtselect/jtonlinesafety/129/129.pdf.

[16] 'Joint Statement – U.S.-European Union Summit Statement' 2021, Superintendent of Documents, Washington, June 2021, https://www.govinfo.gov/app/details/DCPD-202100509/.

[17] Johnson et al., 'Guide to Cyber Threat Information Sharing'.

[18] Ibid.

and are the detailed steps that prescribe how to perform specific tasks.

Furthermore, the concept of FIMI is the nature of the threat discussed in this report and frames the need for a universal defence framework that describes behaviours in consistent and concise ways to support the identification and recording of FIMI activities. To conceptualize the nature of the threat from FIMI, a simple metric of observe, investigate, and identify was applied to an incident in the Operation Ghostwriter campaign. The ABC model,[19] was used to investigate: (A) What kind of actors are involved? (B) What behaviours (activities) are exhibited? and (C) What content is being created and distributed?

[19] Camille François, 'Actors, Behaviors, Content: A Disinformation ABC – Highlighting Three Vectors of Viral Deception to Guide Industry & Regulatory Responses', Graphika and Berkman Klein Center for Internet & Society at Harvard University, September 2019, https://science.house.gov/imo/media/doc/Francois%20Addendum%20to%20Testimony%20-%20ABC_Framework_2019_Sept_2019.pdf.

# 3. DISARM: Framework

### Overview

The DISARM framework grew out of a need to describe disinformation behaviours in consistent, concise ways across groups.[20]

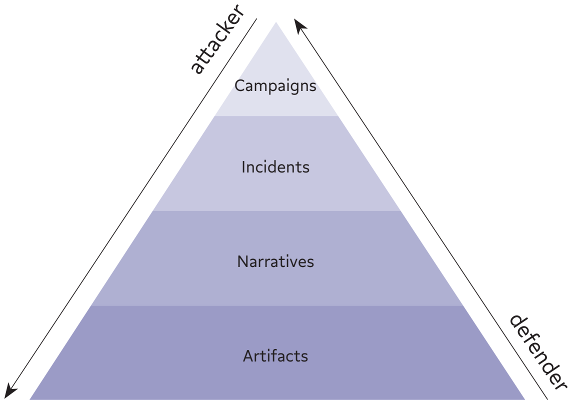
Creating a universal approach to identify and record disinformation activity not only unlocks valuable information within the data, but it also enables concerted action based on the intelligence.

DISARM was created as a universal approach to identify and record disinformation attacks throughout the security community. It is an open-source repository of disinformation tactics, techniques, and procedures as well as counter-responses to attacks. DISARM was created to show how information security principles and practices can be used to increase resilience against and to counter foreign information manipulation and interference activity, specifically disinformation campaigns.[21]

Defence against information manipulation and interference is an ecosystem, and connecting practitioners from different organizations and specialisms was a design priority for DISARM. The Disinformation Pyramid,[22] shown in Figure 1, was created to help the infosec community work together and shift from assessing the problem to being able to meaningfully defend against it.

From top to bottom, it contains the following four layers: campaigns, incidents, narratives, and artifacts. Table 1 includes definitions of each layer, as described during the design phase. The Disinformation Pyramid highlights two perspectives: the creators of disinformation (attackers), who see the whole of the pyramid from top to bottom, and the creators of counter-responses (defenders), who usually see it from the bottom up. The targets of the attack are not included in the diagram.

### Figure 1. The Disinformation Pyramid



[20] DISARM, 'DISARM Design Guide', Github, 2021, p. 2, https://github.com/DISARMFoundation/DISARMframeworks/blob/main/DISARM_DOCUMENTATION/00_AMITT_Design_Guide_version1.pdf.
[21] Ibid.
[22] DISARM, 'DISARM Design Guide', p. 7.

Table 1. Definitions of each layer of the Disinformation Pyramid

| Layer | Title | Definition |
|---|---|---|
| Top | Campaigns | Advanced persistent threats predominantly created by nation-state actors using information manipulation and interference with long-term objectives. They consist of multiple incidents. |
| Second from top | Incidents | Shorter-term sets of information manipulation and interference activity with specific objectives, e.g., to change people's beliefs, emotions, or behaviours. Bursts of activity may be opportunistic, and created by individuals, groups, and organizations. |
| Third from top | Narratives | Stories told to shape beliefs, emotions, and the actions of targeted individuals and groups. |
| Bottom | Artifacts | Messages, images, accounts, relationships, and groups that a malicious actor uses to create narratives and incidents. Artifacts are visible in each incident, often in large volumes, and they form the layer that data scientists and other data specialists usually work on. |

**Design**

The DISARM framework was custom-built and tested through the large-scale collaboration of experts in the information security community. It is intended to be used by that same community, as well as a wider stakeholder group encompassing governments, international organizations and institutions, platforms, academia, private industry, and civil society as a collective defence framework. DISARM was designed based on the MITRE ATT&CK framework and was built to work with existing languages and serialization formats such as the Structured Threat Information Expression (STIX), which is used to exchange cyber-threat intelligence. STIX enables organizations to share cyber-threat intelligence with one another in a consistent and machine-readable manner, allowing security communities to better understand what computer-based attacks they are most likely to see, and to anticipate and/or respond to those attacks faster and more effectively.[23] STIX uses the Trusted Automated Exchange of Intelligence Information (TAXII) as the 'transport mechanism for sharing cyber threat intelligence'. TAXII is a protocol that enables the sharing [of cyber threat] information over HTTPS by defining an application programming interface (API) that aligns with common sharing models.[24] Additionally, DISARM was tested to work with existing threat-sharing formats such as the Malware Information Sharing Platform (MISP), which is one of the open-source tools for sharing incident information and analysis, and was adapted for the specific needs of disinformation.[25]

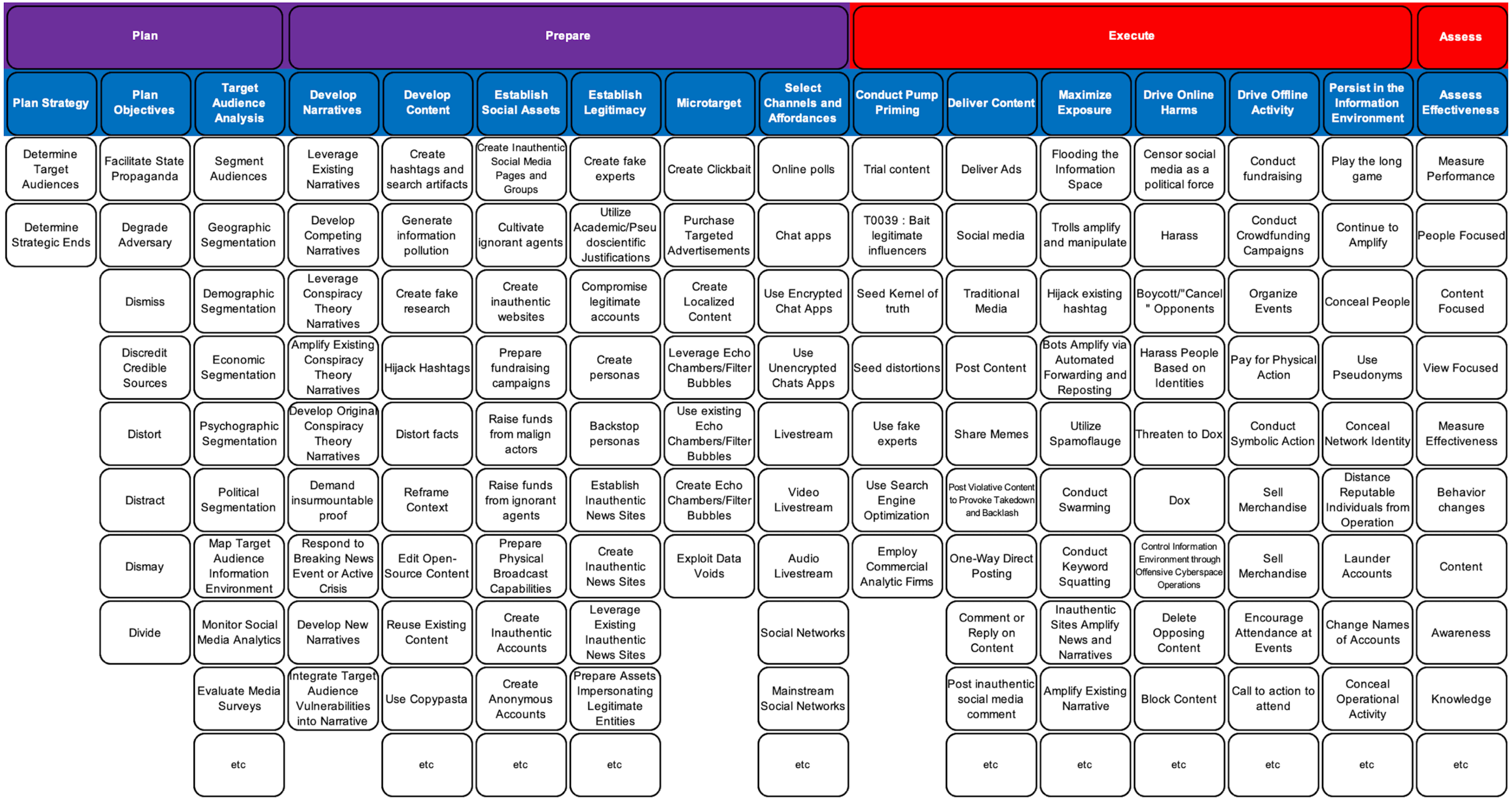
**Frameworks**

Two TTP frameworks were created for DISARM – offence (red – Figure 2) and defence (blue – Figure 3). The individual boxes, known as 'objects', within the frameworks represent tactic stages (which within DISARM specifically mean steps in an incident), and techniques (which

[23] See https://oasis-open.github.io/cti-documentation/.

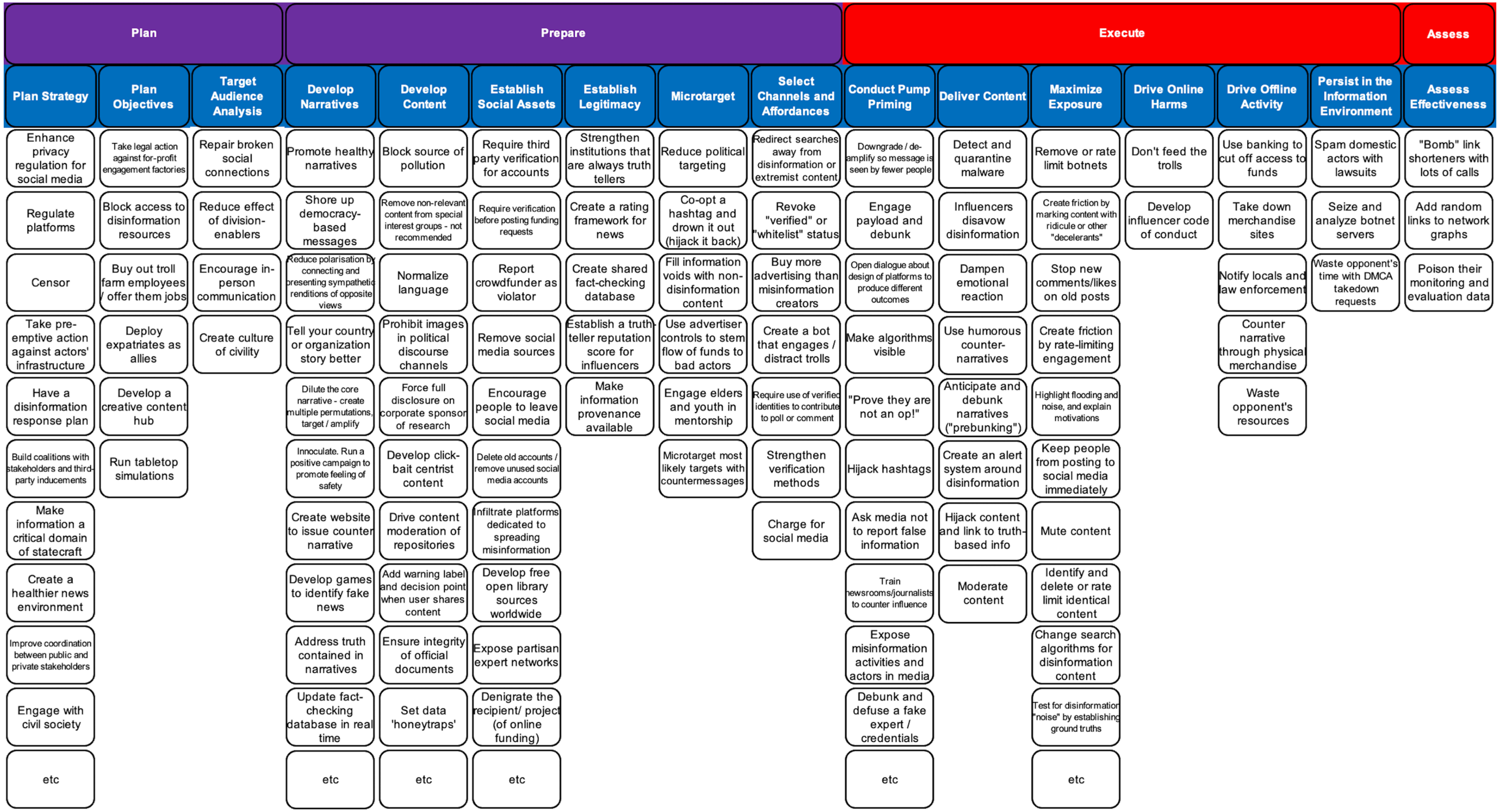
[24] See https://oasis-open.github.io/cti-documentation/.

[25] The Malware Information Sharing Platform (MISP) is an open-source software solution for collecting, storing, distributing, and sharing cyber-security indicators and threats. MISP is one example of the multiple tools available for visualizing such information. DISARM stakeholders also use OpenCTI as a preferred tool for recording, processing, and visualizing disinformation incidents.



| Plan | | | Prepare | | | | | | Execute | | | | | | Assess |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Plan Strategy | Plan Objectives | Target Audience Analysis | Develop Narratives | Develop Content | Establish Social Assets | Establish Legitimacy | Microtarget | Select Channels and Affordances | Conduct Pump Priming | Deliver Content | Maximize Exposure | Drive Online Harms | Drive Offline Activity | Persist in the Information Environment | Assess Effectiveness |
| Determine Target Audiences | Facilitate State Propaganda | Segment Audiences | Leverage Existing Narratives | Create hashtags and search artifacts | Create Inauthentic Social Media Pages and Groups | Create fake experts | Create Clickbait | Online polls | Trial content | Deliver Ads | Flooding the Information Space | Censor social media as a political force | Conduct fundraising | Play the long game | Measure Performance |
| Determine Strategic Ends | Degrade Adversary | Geographic Segmentation | Develop Competing Narratives | Generate information pollution | Cultivate ignorant agents | Utilize Academic/Pseudoscientific Justifications | Purchase Targeted Advertisements | Chat apps | T0039 : Bait legitimate influencers | Social media | Trolls amplify and manipulate | Harass | Conduct Crowdfunding Campaigns | Continue to Amplify | People Focused |
| | Dismiss | Demographic Segmentation | Leverage Conspiracy Theory Narratives | Create fake research | Create inauthentic websites | Compromise legitimate accounts | Create Localized Content | Use Encrypted Chat Apps | Seed Kernel of truth | Traditional Media | Hijack existing hashtag | Boycott/"Cancel" Opponents | Organize Events | Conceal People | Content Focused |
| | Discredit Credible Sources | Economic Segmentation | Amplify Existing Conspiracy Theory Narratives | Hijack Hashtags | Prepare fundraising campaigns | Create personas | Leverage Echo Chambers/Filter Bubbles | Use Unencrypted Chats Apps | Seed distortions | Post Content | Bots Amplify via Automated Forwarding and Reposting | Harass People Based on Identities | Pay for Physical Action | Use Pseudonyms | View Focused |
| | Distort | Psychographic Segmentation | Develop Original Conspiracy Theory Narratives | Distort facts | Raise funds from malign actors | Backstop personas | Use existing Echo Chambers/Filter Bubbles | Livestream | Use fake experts | Share Memes | Utilize Spamoflauge | Threaten to Dox | Conduct Symbolic Action | Conceal Network Identity | Measure Effectiveness |
| | Distract | Political Segmentation | Demand insurmountable proof | Reframe Context | Raise funds from ignorant agents | Establish Inauthentic News Sites | Create Echo Chambers/Filter Bubbles | Video Livestream | Use Search Engine Optimization | Post Violative Content to Provoke Takedown and Backlash | Conduct Swarming | Dox | Sell Merchandise | Distance Reputable Individuals from Operation | Behavior changes |
| | Dismay | Map Target Audience Information Environment | Respond to Breaking News Event or Active Crisis | Edit Open-Source Content | Prepare Physical Broadcast Capabilities | Create Inauthentic News Sites | Exploit Data Voids | Audio Livestream | Employ Commercial Analytic Firms | One-Way Direct Posting | Conduct Keyword Squatting | Control Information Environment through Offensive Cyberspace Operations | Sell Merchandise | Launder Accounts | Content |
| | Divide | Monitor Social Media Analytics | Develop New Narratives | Reuse Existing Content | Create Inauthentic Accounts | Leverage Existing Inauthentic News Sites | | Social Networks | | Comment or Reply on Content | Inauthentic Sites Amplify News and Narratives | Delete Opposing Content | Encourage Attendance at Events | Change Names of Accounts | Awareness |
| | | Evaluate Media Surveys | Integrate Target Audience Vulnerabilities into Narrative | Use Copypasta | Create Anonymous Accounts | Prepare Assets Impersonating Legitimate Entities | | Mainstream Social Networks | | Post inauthentic social media comment | Amplify Existing Narrative | Block Content | Call to action to attend | Conceal Operational Activity | Knowledge |
| | | etc | | etc | etc | etc | | etc | | etc | etc | etc | etc | etc | etc |

Figure 2. Offence [attack] Framework



| Plan | | | Prepare | | | | | | Execute | | | | | | Assess |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Plan Strategy | Plan Objectives | Target Audience Analysis | Develop Narratives | Develop Content | Establish Social Assets | Establish Legitimacy | Microtarget | Select Channels and Affordances | Conduct Pump Priming | Deliver Content | Maximize Exposure | Drive Online Harms | Drive Offline Activity | Persist in the Information Environment | Assess Effectiveness |
| Enhance privacy regulation for social media | Take legal action against for-profit engagement factories | Repair broken social connections | Promote healthy narratives | Block source of pollution | Require third party verification for accounts | Strengthen institutions that are always truth tellers | Reduce political targeting | Redirect searches away from disinformation or extremist content | Downgrade / de-amplify so message is seen by fewer people | Detect and quarantine malware | Remove or rate limit botnets | Don't feed the trolls | Use banking to cut off access to funds | Spam domestic actors with lawsuits | "Bomb" link shorteners with lots of calls |
| Regulate platforms | Block access to disinformation resources | Reduce effect of division-enablers | Shore up democracy-based messages | Remove non-relevant content from special interest groups - not recommended | Require verification before posting funding requests | Create a rating framework for news | Co-opt a hashtag and drown it out (hijack it back) | Revoke "verified" or "whitelist" status | Engage payload and debunk | Influencers disavow disinformation | Create friction by marking content with ridicule or other "decelerants" | Develop influencer code of conduct | Take down merchandise sites | Seize and analyze botnet servers | Add random links to network graphs |
| Censor | Buy out troll farm employees / offer them jobs | Encourage in-person communication | Reduce polarisation by connecting and presenting sympathetic renditions of opposite views | Normalize language | Report crowdfunder as violator | Create shared fact-checking database | Fill information voids with non-disinformation content | Buy more advertising than misinformation creators | Open dialogue about design of platforms to produce different outcomes | Dampen emotional reaction | Stop new comments/likes on old posts | | Notify locals and law enforcement | Waste opponent's time with DMCA takedown requests | Poison their monitoring and evaluation data |
| Take pre-emptive action against actors' infrastructure | Deploy expatriates as allies | Create culture of civility | Tell your country or organization story better | Prohibit images in political discourse channels | Remove social media sources | Establish a truth-teller reputation score for influencers | Use advertiser controls to stem flow of funds to bad actors | Create a bot that engages / distract trolls | Make algorithms visible | Use humorous counter-narratives | Create friction by rate-limiting engagement | | Counter narrative through physical merchandise | | |
| Have a disinformation response plan | Develop a creative content hub | | Dilute the core narrative - create multiple permutations, target / amplify | Force full disclosure on corporate sponsor of research | Encourage people to leave social media | Make information provenance available | Engage elders and youth in mentorship | Require use of verified identities to contribute to poll or comment | "Prove they are not an op!" | Anticipate and debunk narratives ("prebunking") | Highlight flooding and noise, and explain motivations | | Waste opponent's resources | | |
| Build coalitions with stakeholders and third-party inducements | Run tabletop simulations | | Innoculate. Run a positive campaign to promote feeling of safety | Develop click-bait centrist content | Delete old accounts / remove unused social media accounts | | Microtarget most likely targets with countermessages | Strengthen verification methods | Hijack hashtags | Create an alert system around disinformation | Keep people from posting to social media immediately | | | | |
| Make information a critical domain of statecraft | | | Create website to issue counter narrative | Drive content moderation of repositories | Infiltrate platforms dedicated to spreading misinformation | | | Charge for social media | Ask media not to report false information | Hijack content and link to truth-based info | Mute content | | | | |
| Create a healthier news environment | | | Develop games to identify fake news | Add warning label and decision point when user shares content | Develop free open library sources worldwide | | | | Train newsrooms/journalists to counter influence | Moderate content | Identify and delete or rate limit identical content | | | | |
| Improve coordination between public and private stakeholders | | | Address truth contained in narratives | Ensure integrity of official documents | Expose partisan expert networks | | | | Expose misinformation activities and actors in media | | Change search algorithms for disinformation content | | | | |
| Engage with civil society | | | Update fact-checking database in real time | Set data 'honeytraps' | Denigrate the recipient/ project (of online funding) | | | | Debunk and defuse a fake expert / credentials | | Test for disinformation "noise" by establishing ground truths | | | | |
| etc | | | etc | etc | etc | | | | etc | | etc | | | | |

Figure 3. Defence [countermeasures] Framework

Table 2. Relationship of objects by phase

| PHASES | OFFENCE (RED) | DEFENCE (BLUE) |
|---|---|---|
| PLAN | Incident | Response types |
| | Tactics | Actors |
| PREPARE | Techniques | Meta techniques |
| EXECUTE | Tasks | Counters |
| ASSESS | Tactic T0012 | Tactic T0012 |

within DISARM specifically mean activities at each tactic stage). There are also data objects to show how the frameworks are used in practice. All objects in the DISARM framework are cross-referenced to hierarchical symbiotic 'parent-child' relationships, offering a 'one-click' solution to counter-response planning. Furthermore, they allow for the development of attack and counter-simulations, which can accelerate training.

### Objects

Not only do objects form the building blocks of the DISARM framework, but they also enable the creation of offence and defence strategies. While FIMI actors leverage tactics, techniques, and tasks throughout all four phases of an incident, Table 2 indicates the relationship of objects by phase.

- *Phases*: higher-level groupings of tactics
- *Incidents*: incident descriptions used to identify attack frameworks
- *Tactics*: a set of activities/techniques that someone running an information manipulation and interference incident is likely to use
- *Technique*: an activity that might be seen within tactics
- *Tasks*: things that need to be done at each phase
- *Response types*: the course-of-action categories used to create counters
- *Actors*: resources needed to run countermeasures
- *Meta techniques*: a higher-level grouping for countermeasures
- *Counters*: countermeasures to tactics and techniques

### Resources for DISARM

- *Attack framework (red)*: framework for describing disinformation incidents
- *Countermeasures framework (blue)*: framework to defend against disinformation incidents (directly mapped to attack tactics and techniques)
- *Guides*: user guides, design guides, and detailed documentation for all objects
- *History*: documentation on origins, earlier models, and reports
- *Playbooks*: scenarios and simulations for training
- *MISP*: an open-source tool for sharing incident information and analysis, adapted for the specific needs of disinformation[26]
- *Navigator*: the web-based capability development tool and user interface that eliminates the need to work directly with STIX/JSON in the creation of attack (red) and counter (blue) simulations and live incidents. Features include exportable layers in multiple outputs (STIX/JSON code, Excel spreadsheet, PNG image).

[26] MISP is one example of the multiple tools available for visualizing such information.

### DISARM stakeholders and primary users

Comparable to the diverse stakeholder group consulted by the European External Action Service in their aim of developing a shared understanding and coordinated policy response to the threat [of FIMI], [27] and with the need for a whole-of-society approach to increase resilience against and to counter FIMI in mind, the DISARM stakeholders include governments, international organizations and institutions, platforms, academia, private industry, and civil society. The global counter-disinformation community subsist within this group of stakeholders and include strategic communications practitioners and FIMI threat analysts.

### Maintaining DISARM

Since DISARM is an open-source framework with a Creative Commons (CC-BY) licence, all stakeholders and interest groups are able to play a role in keeping it updated. Accordingly, it is housed on the leading developer platform, GitHub.[28]

DISARM Foundation maintains the DISARM framework with a focus on its accessibility and ease of use. It is a funded, non-profit organization that ensures the independence and continuity of the framework. Additionally, reflecting a transatlantic and cross-industry approach, the board of DISARM Foundation comprises US and EU NGOs, and practitioners. The DISARM Foundation is establishing robust governance procedures to enable the growing global community of practitioners to suggest improvements and offer code, as well as help promote and support the framework's wider adoption.

## Summary

DISARM is a well-conceived framework that aims to describe disinformation behaviours in consistent and concise ways and, furthermore, to become the universal approach to identify and record disinformation attacks. Having discussed DISARM's design, frameworks, attributes, stakeholders and users, the case study of Operation Ghostwriter is presented in the next section of this report. This case was identified as the FIMI campaign use case for the evaluation of the DISARM framework. The tactics, techniques, and incidents of Operation Ghostwriter were applied to DISARM (using the framework's coding: object types, tactic stages and techniques) for categorization and encoding.

[27] European External Action Service (EEAS), 'Tackling Disinformation'.

[28] See https://github.com/DISARMFoundation. MisinfosecWG created the original AMITT frameworks. The Red Framework was piloted in December 2018 and was subsequently refined in a Credibility Coalition Misinfosec seminar. Soon after, the Blue Framework arose as a collection of potential disinformation countermeasures at a Coalition Misinfosec seminar that took place in November 2019. (CogSecCollab is a non-profit that spun out of MisinfosecWG.) CogSecCollab was a group of volunteers, acting as an incubator of AMITT and maintaining the AMITT family of models: AMITT-STIX, the AMITT Red Framework (for the creation of disinformation TTPs), and the AMITT Blue Framework (for countermeasures and mitigations). It is worth noting that AMITT has been used in the CTI League's COVID-19 responses and was trialled by NATO and the EEAS.

# 4. Operation Ghostwriter: Poland

## Overview

The issues of scale, potency, and proliferation of disinformation,[29] combined with the power to influence people's ideas and behaviours regarding online information activity,[30] are explored in this section of the report through Operation Ghostwriter, and its influence in Poland. Operation Ghostwriter is used to highlight the pervasive effectiveness of a recent foreign information manipulation and interference campaign, which exemplifies a campaign of considerable scale and potency through its disrupting effect on Polish politics and its prolific use of tactics and techniques.

In July 2020, an intelligence report was published[31] that described a cyber-enabled influence campaign with a threat categorization of UNC1151[32] named "Operation Ghostwriter". Operation Ghostwriter has been attributed to multiple countries and initially targeted audiences in Lithuania, Latvia, and Poland with narratives that were critical of NATO's presence in Eastern Europe. In September 2021, the European Council elevated its status by referring to it by name in the diplomatic environment.

The High Representative of the Union for Foreign Affairs and Security Policy claimed: "Some EU Member States have observed malicious cyber activities, collectively designated as Operation Ghostwriter… Such activities are unacceptable as they seek to threaten our integrity and security, democratic values and principles and the core functioning of our democracies…"[33]

There is reason to believe that Operation Ghostwriter did not restrict its operations to Lithuania, Latvia, and Poland – that UNC1151 targeted entities outside of the Baltics, Poland, Ukraine, and Germany.[34] Hardly a new phenomenon, this extensive operation was designed to disseminate disinformation across Eastern Europe and the Baltics (and possibly further

[29] Hadley Newman, 'Understanding The Differences Between Disinformation, Misinformation, Malinformation and Information – Presenting The DMMI Matrix', evidence to UK Government Joint Committee for the Draft Online Safety Bill, 2020, https://publications.parliament.uk/pa/jt5802/jtselect/jtonlinesafety/129/129.pdf.

[30] S. M Diao et al., 'A novel opinion dynamics model based on expanded observation ranges and individuals' social influences in social networks', *Physica A: Statistical Mechanics and its Applications*, volume 415 (2014): 220–228.

[31] Mandiant, '"Ghostwriter" Influence Campaign: Unknown Actors Leverage Website Compromises and Fabricated Content to Push Narratives Aligned with Russian Security Interest', FireEye Blog, 2020, https://www.fireeye.com/content/dam/fireeye-www/blog/pdfs/Ghostwriter-Influence-Campaign.pdf. Mandiant is a publicly traded American cybersecurity firm. In December 2013, it was acquired by FireEye and in June 2021, Mandiant sold the FireEye product line, name, and roughly 1,300 employees to Symphony Technology Group. For this reason, Mandiant reports are often held on the FireEye domain.

[32] Mandiant has three types of threat categorizations, their classification letters, and associated characteristics. For its part, UNC1151 is classified as an 'uncategorized threat' (rather than 'advanced persistent' or 'financially motivated') and derives its name from where it falls on Mandiant's list.

[33] EU Council, 'Declaration by the High Representative on behalf of the European Union on respect for the EU's democratic processes', 2021, https://www.consilium.europa.eu/en/press/press-releases/2021/09/24/declaration-by-the-high-representative-on-behalf-of-the-european-union-on-respect-for-the-eu-s-democratic-processes/.

[34] Prevailion, 'Diving Deep into UNC1151's Infrastructure: Ghostwriter and beyond', 2021, https://www.prevailion.com/diving-deep-into-unc1151s-infrastructure-ghostwriter-and-beyond/.

afield) and appears to have been in place for at least the past five years. The full scope and reach of its operations were not well known, lacked delineation, and remain largely underreported.

### Geographical focus and TTPs

Within Poland, Ghostwriter became prominent with five campaigns that "took place between October 2020 and January 2021 in which the social media accounts of Polish officials were compromised and used to disseminate narratives intended to discredit the Polish government and to widen existing domestic political divisions".[35]

Over the course of the campaign, the narratives, targeting, tactics, techniques, and procedures (TTPs) grew in scale and influence. The distinct activities that were waged against Poland have been collated and are presented in Table 3. The techniques with the corresponding incidents and the date of occurrence are detailed. In this table, the techniques have been divided by content and channels. Some consistency in techniques that were employed by the actor across multiple incidents can be observed.

Row 1 highlights a disinformation incident targeting Poland. The incident suggested that a Lithuanian nuclear plant was leaking radioactive waste, threatening the lives and well-being of Lithuanians in its vicinity, and endangering Poles who were living and working near the border. Using simplistic techniques, this incident was based on fabricated statements that were first disseminated through websites that were compromised via phishing efforts and social media accounts, and then through a fake government website. Table 3 lists 14 other examples of disinformation incidents that Operation Ghostwriter orchestrated against Poland.

### Identified content techniques

In Row 1 (Table 3), the example of the radioactive waste leak disinformation incident included 'fabricated statement', but single content techniques were not the norm. As Table 5 illustrates, fabricated statements were typically used up to four times, and most incidents employed more than one technique. Row 2 (Table 3) gives the example of a salacious incident, where Polish, Lithuanian, and American officials were claimed to have been involved in a military prostitution scandal.

As Mandiant, the American cybersecurity firm, notes: "The operation promoted a narrative alleging that the Polish Ministry of National Defence uses female officers to provide 'escort services' for important Polish and foreign officials. A named female officer was alleged to have provided such services for Polish President Andrzej Duda, Lithuanian Foreign Minister Gabrielius Landsbergis and senior U.S. military representatives."[36] In the wake of this fabricated scandal, the Polish Ministry of Science and Higher Education[37] sent a tweet confirming that this was a disinformation cyberattack. At the same time, Stanisław Żaryn, Ministry spokesperson, tweeted: "The intention of described

[35] Mandiant, 'Ghostwriter Update: Cyber Espionage Group UNC1151 Likely Conducts Ghostwriter Influence Activity', 2021, p. 3, https://content.fireeye.com/web-assets/rpt-unc1151-ghostwriter-update.

[36] Mandiant, 'Ghostwriter Update', p. 16.

[37] Ministerstwo Edukacji i Nauki, Twitter, 2021, https://twitter.com/MEIN_GOV_PL/status/1364895680284815362.

## Table 3. Operation Ghostwriter activity against Poland 2020–2021[38]

| | Techniques | | Incident | Date |
|---|---|---|---|---|
| | **Content** | **Channels** | | |
| 1 | Fabricated statements | Compromised website and social media accounts, fake government website | Radioactive waste leaked from Lithuanian nuclear plant poses danger to Poles living near border | 2021<br>17/03 |
| 2 | Fabricated article and statements, falsified social media posts, compromising photos of a sexual nature | Compromised website and social media accounts, spoof emails | Polish, Lithuanian, and American officials involved in military prostitution scandal | 2021<br>25/02 |
| 3 | Compromising photos of a sexual nature, fabricated article, and social media posts, impersonating a known journalist | Compromised social media account | Polish politician posts compromising sexual photos of former PiS[39] mayoral candidate | 2021<br>18/01 |
| 4 | Falsified social media post | Compromised social media account | PiS is the party of "Murderers, Thieves, and Executioners" | 2021<br>12/01 |
| 5 | Falsified social media post | Compromised social media account | Polish minister condemns female activists, uses racial slurs | 2020<br>15/12 |
| 6 | Fabricated press release, statement, and op-edv | Compromised website and social media account, fake email | Polish diplomat arrested entering Lithuania; Lithuanian conscripts called up for duty; Šiauliai Airport modernization benefits NATO, harms locals | 2020<br>09/12 |
| 7 | Fabricated statement | Fake government website, compromised social media account | Poland trained extremists to destabilize Lithuania | 2020<br>27/11 |
| 8 | Compromising social media post and explicit photos that were sexual in nature | Compromised social media account | Polish MP brags about new female secretary | 2020<br>19/11 |
| 9 | Fabricated social media post | Compromised social media account | Polish MP calls pro-choice activists "drug addicts, prostitutes and child killers" | 2020<br>29/10 |
| 10 | Fabricated article | Compromised social media account and website | NATO preparing for war with Russia on Polish, Latvian and Lithuanian soil | 2020<br>22/10 |
| 11 | Fabricated blog, impersonating NATO unit | Email spoofing | Lithuanian military officer arrested in Poland for espionage | 2020<br>21/07 |

[38] Adapted from Mandiant, 'Ghostwriter Update', p. 14.

[39] PiS (*Prawo i Spraw*iedliwość – Law and Justice) is a right-wing populist and national-conservative political party in Poland.

| | Techniques | | Incident | Date |
|---|---|---|---|---|
| | **Content** | **Channels** | | |
| 12 | Falsified interview transcripts, fabricated quotes | Compromised website | Commanding general of US Army in Europe criticizes Polish, Baltic militaries | 2020<br>27/05 |
| 13 | Fabricated correspondence | Compromised website and email spoofing | Polish soldiers should rebel against American "Occupational Forces" | 2020<br>22/04 |
| 14 | Falsified quote | - | US relocated nuclear weapons from Turkey to Germany, Poland, Baltics | 2020<br>21/02 |
| 15 | Falsified quote | - | USARMEUR Chief of Staff criticized Polish military | 2020<br>18/02 |

## Table 4. Techniques and their occurrence

| **Content technique** | **Occurrence** |
|---|---|
| Falsified social media post | 6 |
| Fabricated statement | 4 |
| Compromising sexual photos | 3 |
| Fabricated article | 3 |
| Fabricated quotes | 3 |
| Fabricated blog impersonating NATO unit | 1 |
| Fabricated correspondence | 1 |
| Fabricated op-ed | 1 |
| Fabricated press release | 1 |
| Falsified interview transcripts | 1 |

## Table 5. Channel techniques and their occurrence

| **Channel technique** | **Occurrence** |
|---|---|
| Compromised social media account | 10 |
| Compromised website | 6 |
| Spoofed email | 4 |
| Spoofed government website | 2 |

information activities is to hit the image of the President of Poland, the Polish Army, and to create tensions between PL [Poland] and LT [Lithuania]. The indicated lines of influence correspond with the Russian information offensive against Poland, Lithuania, and NATO."[40]

Meanwhile, on February 26, 2021, the Lithuanian Ministry of Foreign Affairs published an official statement confirming an "information attack on Lithuanian-Polish relations", citing the individuals who were implicated in the incident.[41]

In Row 2, Table 3, four interrelated content techniques were employed, namely, 'fabricated statements'; 'falsified social media posts'; 'fabricated article'; and the creation of 'compromising sexual photos' of the Polish, Lithuanian, and American officials. As Table 4 illustrates, 'falsified social media posts' were the most common content techniques used by UNC1151 in the Operation Ghostwriter incidents that targeted Poland (6 times), followed by 'fabricated statement' (4 times), and then 'compromising sexual photos', 'fabricated article', and 'fabricated quotes' (3 times each).

## Identified channel techniques

It is worth noting that the techniques used to implicate Polish, Lithuanian, and American Officials in a military prostitution scandal were both sophisticated and complex. As demonstrated in Row 2 (Table 3), the channel techniques consisted of compromising authoritative websites and social media accounts, in conjunction with sending emails from spoofed accounts. Mandiant noted that "this dissemination strategy may have been designed to imitate a pattern of official statements and media responses that would unfold around the revelation of an actual government scandal to impart a greater sense of legitimacy to the narrative and potentially increase its reach".[42] This observation illustrates that Ghostwriter operated with a considerable level of sophistication and complexity that comprised different techniques to disseminate the disinformation that had been created by UNC1151. As the spokesperson from the Polish government observed, this incident was "aimed at hitting the credibility of the Polish Armed Forces, as well as ridiculing the most important officials in Poland and creating tensions between Poland and Lithuania. The actions of the info-aggressor follow a scenario known from previous attempts".[43] Clearly, the Polish government did not consider this to be a one-off event but rather another incident in an extensive campaign.

Several channel techniques are common across UNC1151 incidents, as indicated in Table 5. Of them, 'compromised social media accounts' were the most common technique (occurring on at least 10 occasions). To be clear, 10 is the number of occurrences, not the number of individual social media accounts that may have been compromised. Mandiant estimates that number to be considerably higher: "multiple

[40] Stanisław Żaryn, Tweet thread – series of five, Twitter, 2021, https://twitter.com/StZaryn/status/1365339366722265089.

[41] Lietuvos Respublikos Užsienio Reikalų Ministerija, 'Dėl informacinės atakos prieš Lietuvos-Lenkijos santykius', 2021, https://www.urm.lt/default/lt/naujienos/del-informacines-atakos-pries-lietuvos-lenkijos-santykius.

[42] Mandiant, 'Ghostwriter Update', p. 16.

[43] Stanisław Żaryn, Twitter, 2021, https://twitter.com/StZaryn/status/1365339375991685125.

suspected compromised social media accounts belonging to individuals affiliated with political parties in Poland's ruling United Right coalition were used for narrative dissemination."[44]

This exploration of Operation Ghostwriter, and its influence in Poland, provides an example of a potent, prolific, and recent foreign information manipulation and interference campaign delivered at scale that is employed in this report to evaluate DISARM in the definition and cataloguing of disinformation incidents, specifically tactics and techniques. This is needed to fully understand how incidents can be defended, particularly by those who have limited [DISARM] experience.

[44] Mandiant, 'Ghostwriter Update', p. 18.

# 5. DISARM Evaluation

## Overview

### Purpose

In order to develop shared data standards to collectively assess cases of foreign information manipulation and interference, there is a need to test for rapid adoption and ease of use by strategic communications practitioners with no previous DISARM experience; and, in a separate evaluation, consider DISARM's credibility as a universal approach to accurately catalogue and plan responses to disinformation threats and attacks from the perspective of FIMI threat analysts.

### Parameters

**Primary:** DISARM framework and navigator capability development tool (web application on GitHub). *Important*: The DISARM User Guide was neither reviewed nor used throughout the assessment to ensure fair evaluation of intuitiveness for a new user.

**Secondary:** Compatibility with the STIX language and serialization format for codifying incidents, and presence of complementary learning and support tools for rapid adoption (e.g., countermeasures web application for scenario training, user guides, and playbooks).

**Outside of scope:** Accessibility, real-time simulation, and multiple languages were not tested.

### Evaluation steps

The evaluation considered the full DISARM ecosystem of tools (at the time of writing) that a FIMI analyst or strategic communication practitioner might use including the DISARM Framework itself, the DISARM GitHub Repositories that maintain the framework and tools, the DISARM Framework Navigator, and the DISARM STIX implementation.

**Pre-evaluation:** Desk research and semi-structured interviews were conducted with the DISARM co-creators to establish context and familiarity with DISARM and the environment it was built to serve, and to identify potential gaps (there were none) that would prevent a useful assessment.

**Usability evaluation:** A simulation was conducted using Operation Ghostwriter, the defined user capability personas, and the DISARM framework. Next, a sample of techniques from the case study was used to plan countermeasures using the DISARM navigator web application, and the incident was manually codified using STIX. The end-to-end process was evaluated using a heuristic assessment model.

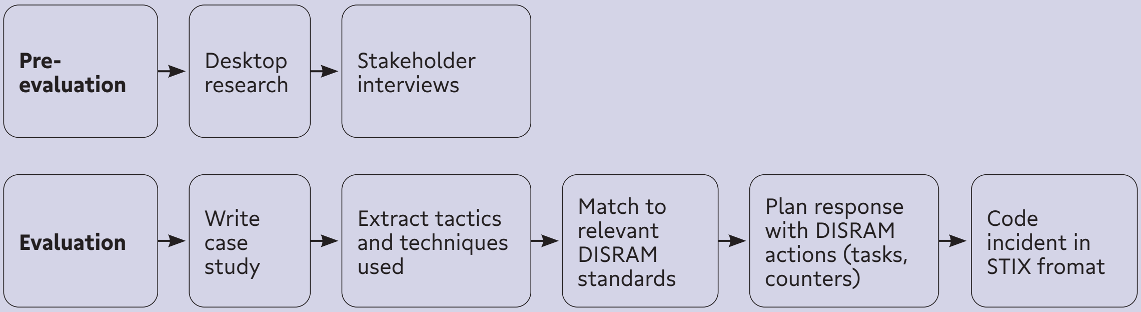
**Credibility evaluation:** Existing and widely adopted principles designed to assess credibility of standardized frameworks were identified and used to assess DISARM.

## Methodology

Following the desk research and interviews (pre-evaluation), the evaluation comprised several actions including the writing of a case study, the extraction of techniques from the case study, matching them to relevant DISARM objects, planning the response with DISARM (tasks, countermeasures) and coding incidents in STIX language and serialization format. The steps are presented in Figure 4.

The experience of using DISARM in the test scenarios was validated against the Heuristics

Figure 4. Steps in pre-evaluation and evaluation


Pre-evaluation
Desktop research
Stakeholder interviews
Evaluation
Write case study
Extract tactics and techniques used
Match to relevant DISRAM standards
Plan response with DISRAM actions (tasks, counters)
Code incident in STIX fromat

Model for evaluating usability, which was suitable for maximum explanatory capability of the end-to-end process. In this phase, the credibility of DISARM – through the perspective of FIMI threat analysts – was validated against an existing and widely adopted set of principles designed to assess the credibility of standardized frameworks.

## Course of action

### Interviews

For the pre-evaluation phase of the assessment, semi-structured interviews were conducted with the DISARM co-creators. They discussed the history of the product, including the work they delivered to the United Nations Development Programme and World Health Organization using DISARM, and their vision for its use.

### Case study

The purpose of conducting the case study was to identify the techniques to apply to the DISARM framework for categorization (using DISARM's coding: object types, tactic stages and techniques) and encoding in the STIX language and serialization format. Operation Ghostwriter was a cyber-enabled foreign information manipulation and interference campaign conducted with 'technical support' from UNC1151, and was selected as the subject of the case study. Operation Ghostwriter targeted audiences in Lithuania, Latvia, and Poland with narratives critical of NATO's presence in Eastern Europe. Exploratory in nature, the case study focuses on the disinformation activity that targeted Poland in 2020–2021 and identified the 15 incidents that, with the identified techniques, answered the research question: **What are the instances of UNC1151 using FIMI activity to target Poland?**

### User capability personas

To establish a benchmark for the general knowledge and skills of a strategic communications practitioner and, separately, a FIMI threat analyst – irrespective of their age, education, geography, and gender – "user capability personas" were designed to identify traits based on skills and knowledge and tasks. This enabled the assessment of DISARM's readiness and how easily it could be adopted. Common to both personas was the fact that a TTP's identification methodology is critical to make the data, analysis, and report applicable across different stakeholders and/or institutions.

*The strategic communications practitioner*

The strategic communications practitioner persona was defined as a professional with no previous knowledge or awareness of DISARM, but with the ability to identify fundamental communications tasks such as audience segmentation, narrative creation, content development, and social media paid promotions.

*The FIMI threat analyst*

The FIMI threat analyst was defined as a professional that identifies, catalogues, and responds – encompassing the collection of missing data and reporting – to information manipulation and interference campaigns. The analyst must exchange the information about the disinformation cases with other stakeholders as part of this work. Finally, the findings and conclusions (analysis results) must be visualized, presentable, and easy to understand (explainable).

**Heuristic assessment model**

'Usability Heuristics'[45] was selected as a suitable model (see Annex I). Heuristic models enable the evaluation of intuitive learning capabilities, ongoing interaction and support, and subjective user satisfaction (considering user expectations and experience).[46]

When information is incomplete, heuristics – mental shortcuts – result in answers that satisfy the immediate need but are not necessarily the same, or even as accurate, if probability or logic had been applied.[47] Signals, it is theorized, can be processed heuristically or systematically. The "heuristic-systematic model of information processing" attempts to explain how messages are received, interpreted, and processed.[48] Essentially, people tend to avoid systematic thinking and rely on heuristics[49] when they deem the issue to be of no personal importance or to not have a significant impact on themselves.

This is connected to satisficing (satisfy and suffice), or to what happens when an optimal decision cannot be reached: "decision makers… [try to find] optimum solutions for a simplified world, or… satisfactory solutions for a more realistic world. Neither approach, in general, dominates the other, and both have continued to co-exist in the world of management science."[50] Departing from systematic decision-making, heuristics depends on memories[51] and minimal cognitive effort.[52] It is governed by:

- Availability – already familiar and remembered
- Accessibility – ability to retrieve the memory
- Applicability – relevance and relatability to the decision-making task

[45] J. Nielsen & R. Molich, 'Heuristic evaluation of user interfaces', *CHI '90*, (1990).

[46] J. Nielsen, 'Heuristic Evaluation', in *Usability Inspection Methods*, ed. J. Nielsen and R. L. Mack (New York: John Wiley and Sons, 1994), 25–62.

[47] R. W. Scholz, *Decision Making under Uncertainty: Cognitive Decision Research, Social Interaction, Development and Epistemology* (New York: Elsevier, 1983).

[48] S. Chaiken, 'Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion', *Journal of Personality & Social Psychology*, Volume 39, Issue 5 (1980): 752–766.

[49] S. Chaiken et al., 'Heuristic and systematic processing within and beyond the persuasion context', in *Unintended Thought*, ed. J. S. Veleman and J. A. Bargh (New York: Guilford, 1989), 212–252.

[50] H. A. Simon, 'Rational decision making in business organizations', *The American Economic Review*, Volume 69, Issue 4 (1979): 493–513.

[51] S. Chen et al., 'Motivated Heuristic and Systematic Processing', *Psychological Inquiry*, Volume 10, Issue 1 (1999): 44–57.

[52] Chaiken, 'Heuristic Versus Systematic Information Processing'.

Those who use heuristics are more likely to agree with messages that are presented by experts or advocated by others, and more so when they have not entirely reflected on the content.[53]

This evaluation of DISARM was underpinned by the understanding derived from the heuristic-systematic model of information processing, and implemented by applying the usability heuristic model. Combining the theoretical underpinning and application of an established assessment model facilitated the evaluation of DISARM by assessing the ease of learning and application of the framework, based on user satisfaction (expectations and experience), for the rapid adoption of DISARM.

### Credibility assessment model

'ISEAL Credibility Principles'[54] were selected as suitable criteria (see Annex II). Developed in 2013 and updated in 2021 for the sustainability reporting environment, the principles "help organisations developing standards and similar sustainability tools to understand which attributes of their system are critical to the credibility of their approach, and why this matters for improving sustainability performance and delivering impacts".[55] Although not a sustainability tool, DISARM is a universal approach in a nascent specialist field and therefore is compatible with and can legitimately be evaluated by the ISEAL Credibility Principles.

### Limitations

This report contributes to a growing body of work on FIMI defence and provides practitioners with an understanding of how DISARM could be readily applied to their workflow, and work in harmony with the broader information security ecosystem. There are three caveats worth highlighting:

**Frameworks are a universal language.** While the test was conducted in English, the frameworks were developed based on fundamental communications practice. Therefore, the frameworks would be expected to perform the same way in other languages. It is assumed that sighting, responding, and sharing information about potential and confirmed incidents in multiple languages would be subject to the same challenges and considerations of any multilateral action.

**Responding in real time requires fit-for-purpose application software.** The simulation was conducted retrospectively, using a case study about a known incident (Operation Ghostwriter), and was augmented by secondary research. From this foundation, a recommendation has been made to engage with private industry to develop application software to record, process, and visualize disinformation activity in accordance with the DISARM framework.

[53] A. H. Eagly and S. Chaiken, 'Process theories of attitude formation and change: The elaboration likelihood and heuristic-systematic models', in *The psychology of attitudes*, ed. A. H. Eagly and S. Chaiken (Orlando: Harcourt Brace, 1993), 303–350.

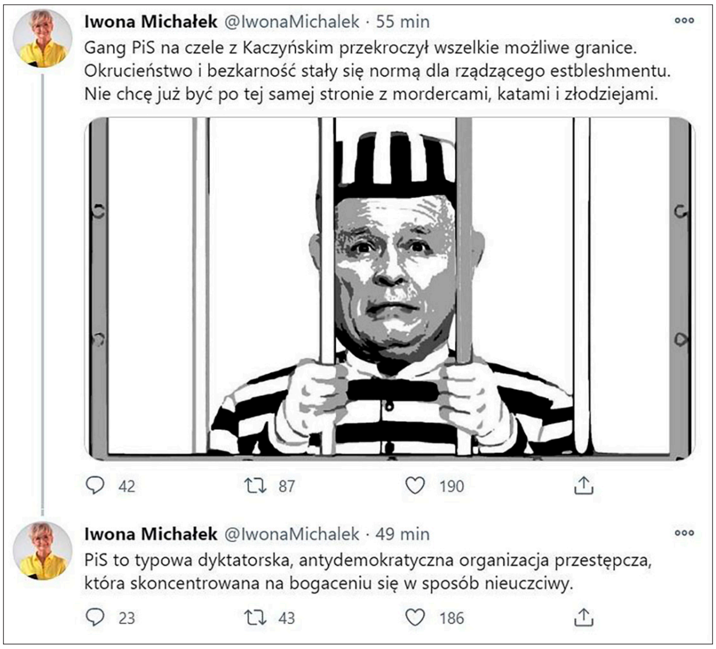
[54] ISEAL Alliance, 'ISEAL Credibility Principles', 2021, https://www.isealalliance.org/sites/default/files/resource/2021–06/ISEAL-Credibility-Principles-V2–2021_EN_ISEAL_June-21.pdf.

[55] See https://www.isealalliance.org/defining-credible-practice/iseal-credibility-principles.

## Table 6. Sighting scenario evaluation process

| **Observe** | **Investigate** | **Identify techniques and tasks** |
|---|---|---|
| *Broad description of what we can see and when* | *Questions to ask to understand what we're seeing* | *Universal description for response, sharing, and analyses (DISARM)* |
| 12 Jan 2021<br>A tweet published to the Twitter account of Iwona Michałek, Poland's deputy minister of development, labour, and technology, disseminated the false narrative that she no longer wanted to be affiliated with the PiS party. The tweet condemned PiS as the party of "murderers, thieves, and executioners" and featured a cartoon of PiS leader Jarosław Kaczyński in prison. See figure 5. | A. What kinds of actors are involved? | Name: Iwona Michałek<br>Description: Poland's deputy minister of development, labour, and technology |
| | B. What **behaviours** (activities) are exhibited? | Technique T0011: Hijack legitimate account<br>Technique T0021: Memes<br>Technique T0038: Twitter (account: @IwonaMichałek) |
| | C. What kind of **content** is being created and distributed? | Task TK0017: content creation<br>Description: false narrative that she no longer wanted to be affiliated with the PiS party. The tweet condemned PiS as the party of "murderers, thieves, and executioners" and featured a cartoon of PiS leader Jarosław Kaczyński in prison. |

## Figure 5. Tweet published to the account of Iwona Michałek



**Iwona Michałek** @IwonaMichalek · 55 min

Gang PiS na czele z Kaczyńskim przekroczył wszelkie możliwe granice. Okrucieństwo i bezkarność stały się normą dla rządzącego estbleshmentu. Nie chcę już być po tej samej stronie z mordercami, katami i złodziejami.

**Iwona Michałek** @IwonaMichalek · 49 min

PiS to typowa dyktatorska, antydemokratyczna organizacja przestępcza, która skoncentrowana na bogaceniu się w sposób nieuczciwy.

**Collective action requires a *collective*.** This assessment was designed to test the speed and ease of learning and applying the framework for the rapid adoption of DISARM. With 4.62 billion social media users (58.4% of the global population) and a 10% year-on-year increase,[56] the size and scale of global and hyper-connected online users presents a vast landscape for FIMI that is too large and complex for a single entity to resolve. An interconnected whole-of-society approach, leveraging the varied capabilities and competences that lie within governments, civil society, and private industry, is required to increase resilience against and to counter FIMI.

## Evaluation

### Usability

Three scenarios were constructed to evaluate whether DISARM made it easy for practitioners to sight, respond to, and analyse disinformation and thus support defence.

1. **Sighting:** use DISARM to understand whether an incident has occurred/is occurring
2. **Response:** use DISARM to prepare a counterattack
3. **Analysis:** use DISARM to share data and measure impact

*Sighting*

The purpose of this scenario was to evaluate whether DISARM made it easy for practitioners to 'sight' incidents and campaigns. For this, a simple metric – *observe, investigate, and identify* – was created and then applied to an incident in the Operation Ghostwriter campaign. The ABC model[57] was used to investigate: (A) What kind of actors were involved? (B) What behaviours (activities) were exhibited? and (C) What content was created and distributed?

The process is outlined in Table 6, where column one contains a description of the incident (observation), column two includes the questions asked to understand what was observed (investigate), and column three details the data extracted by the investigation and codified to DISARM and STIX language and serialization format (identify techniques and tasks).

In the sighting scenario, the navigator web application to create a DISARM framework was labelled "murderers, thieves, and executioners". A screenshot of the framework appears in Figure 6.

The navigator web application was built on MITRE ATT&CK© Navigator and adapted for the DISARM framework to allow practitioners to visualize defensive coverage with a 'point-and-click'. It is also easy to annotate and add custom metadata fields, as shown in Figure 7.

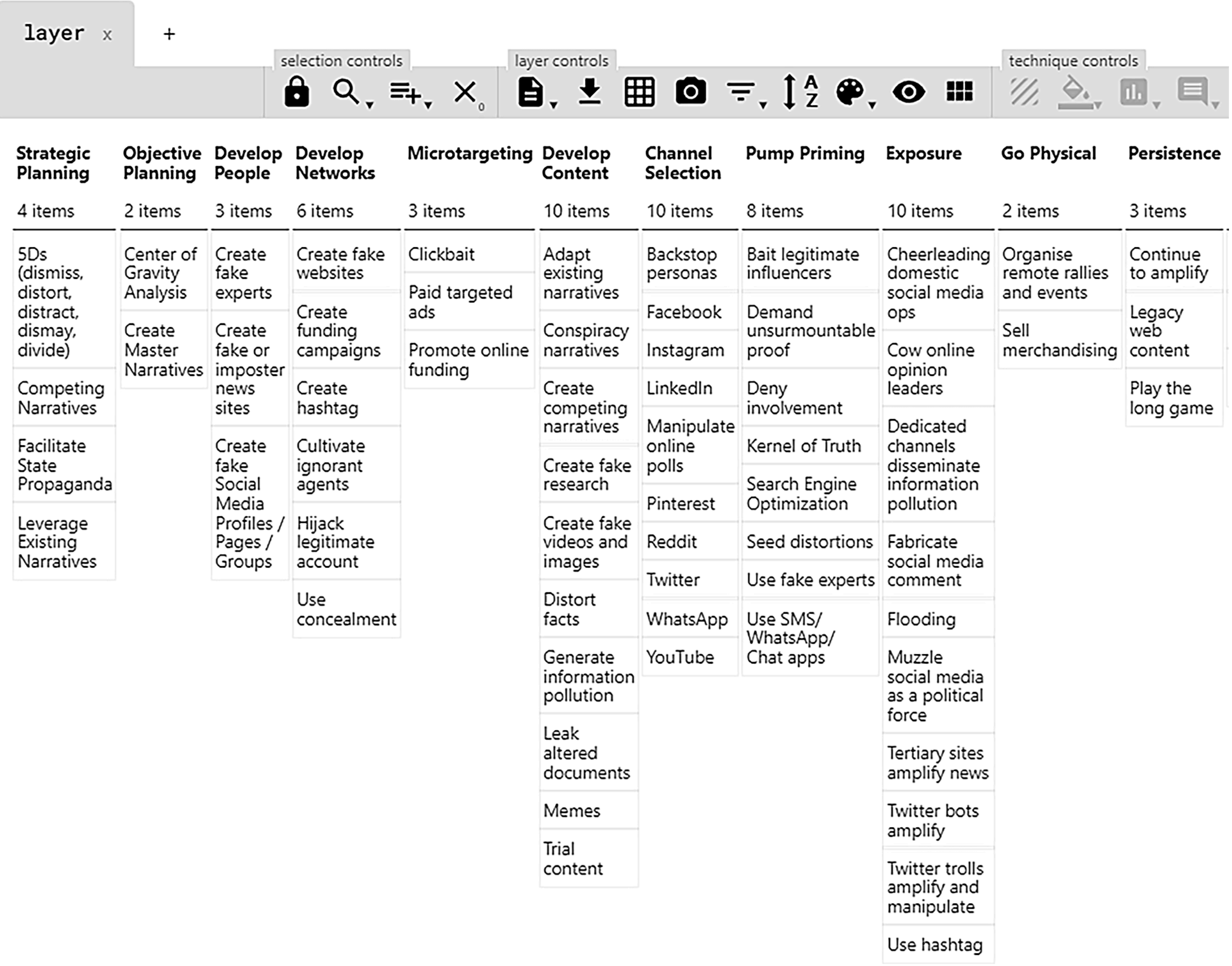
The navigator capability development tools also make it quick and easy to add comments to individual techniques, assign custom colours, and multi-select objects using ctrl and shift commands. Users can access options from the icon-based menu and then right click on the drop-down menu. Hovering over a technique that contains a comment displays the comment as a tool tip/popup window (as shown in the screenshot in Figure 8).

[56] We are Social and Hootsuite, 'Digital 2022 Global Overview Report', January 2022, slide 87, https://www.slideshare.net/DataReportal/digital-2022-global-overview-report-january-2022-v05.

[57] François, 'Actors, Behaviors, Content'.

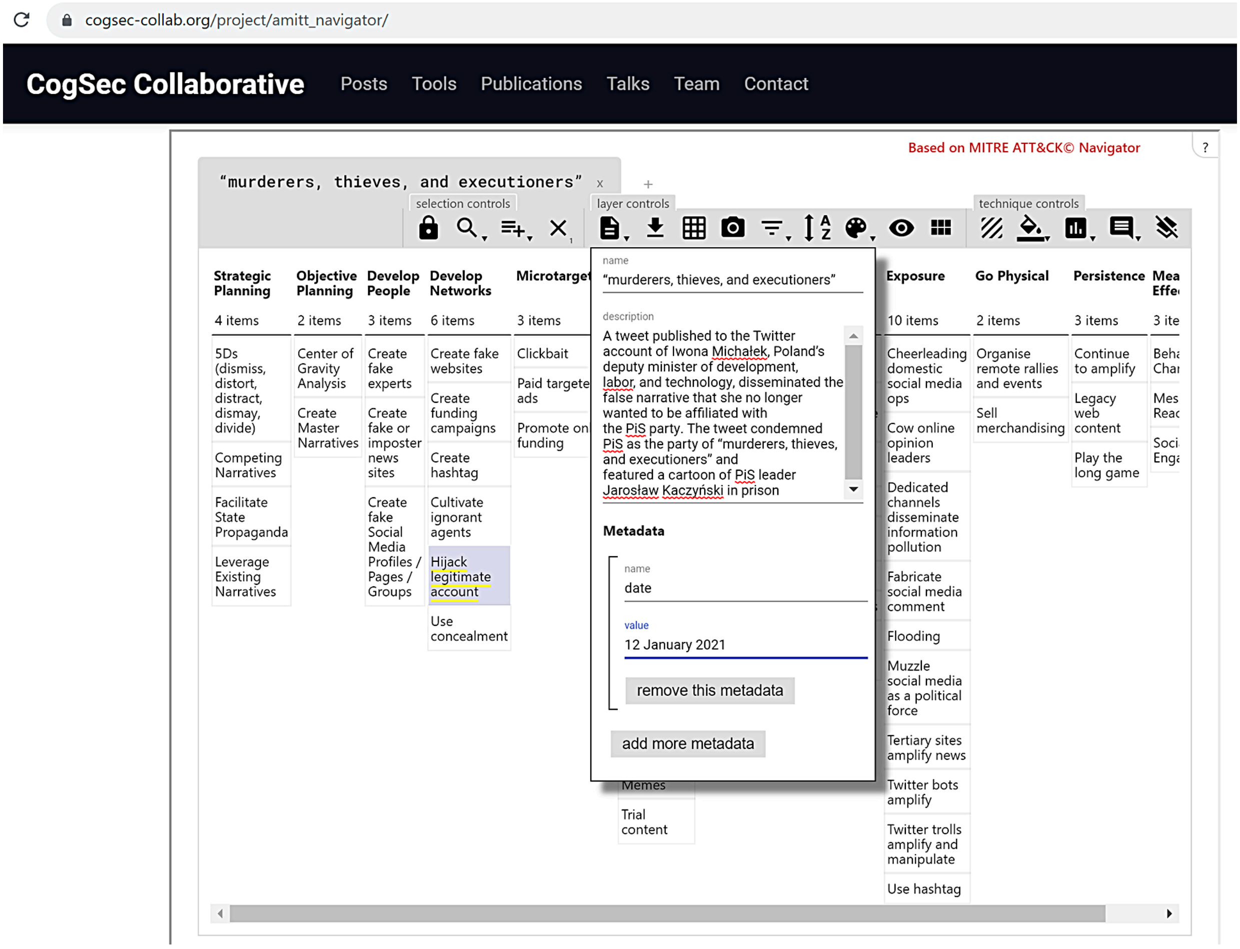
Figure 6. Create DISARM framework using navigator web application

Based on MITRE ATT&CK© Navigator



| Strategic Planning | Objective Planning | Develop People | Develop Networks | Microtargeting | Develop Content | Channel Selection | Pump Priming | Exposure | Go Physical | Persistence |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 items | 2 items | 3 items | 6 items | 3 items | 10 items | 10 items | 8 items | 10 items | 2 items | 3 items |
| 5Ds (dismiss, distort, distract, dismay, divide) | Center of Gravity Analysis | Create fake experts | Create fake websites | Clickbait | Adapt existing narratives | Backstop personas | Bait legitimate influencers | Cheerleading domestic social media ops | Organise remote rallies and events | Continue to amplify |
| Competing Narratives | Create Master Narratives | Create fake or imposter news sites | Create funding campaigns | Paid targeted ads | Conspiracy narratives | Facebook | Demand unsurmountable proof | Cow online opinion leaders | Sell merchandising | Legacy web content |
| Facilitate State Propaganda | | Create fake Social Media Profiles / Pages / Groups | Create hashtag | Promote online funding | Create competing narratives | Instagram | Deny involvement | Dedicated channels disseminate information pollution | | Play the long game |
| Leverage Existing Narratives | | | Cultivate ignorant agents | | Create fake research | LinkedIn | Kernel of Truth | Fabricate social media comment | | |
| | | | Hijack legitimate account | | Create fake videos and images | Manipulate online polls | Search Engine Optimization | Flooding | | |
| | | | Use concealment | | Distort facts | Pinterest | Seed distortions | Muzzle social media as a political force | | |
| | | | | | Generate information pollution | Reddit | Use fake experts | Tertiary sites amplify news | | |
| | | | | | Leak altered documents | Twitter | Use SMS/ WhatsApp/ Chat apps | Twitter bots amplify | | |
| | | | | | Memes | WhatsApp | | Twitter trolls amplify and manipulate | | |
| | | | | | Trial content | YouTube | | Use hashtag | | |

Figure 7. Annotation and addition of custom metadata fields



cogsec-collab.org/project/amitt_navigator/

CogSec Collaborative

“murderers, thieves, and executioners”

| Strategic Planning | Objective Planning | Develop People | Develop Networks | Microtarget |  | Exposure | Go Physical | Persistence | Mea Effe |
|---|---|---|---|---|---|---|---|---|---|
| 4 items | 2 items | 3 items | 6 items | 3 items |  | 10 items | 2 items | 3 items | 3 ite |
| 5Ds (dismiss, distort, distract, dismay, divide) | Center of Gravity Analysis | Create fake experts | Create fake websites | Clickbait |  | Cheerleading domestic social media ops | Organise remote rallies and events | Continue to amplify | Beha Char |
| Competing Narratives | Create Master Narratives | Create fake or imposter news sites | Create funding campaigns | Paid targete ads |  | Cow online opinion leaders | Sell merchandising | Legacy web content | Mes Reac |
| Facilitate State Propaganda |  | Create fake Social Media Profiles / Pages / Groups | Create hashtag | Promote on funding |  | Dedicated channels disseminate information pollution |  | Play the long game | Soci Enga |
| Leverage Existing Narratives |  |  | Cultivate ignorant agents |  |  | Fabricate social media comment |  |  |  |
|  |  |  | Hijack legitimate account |  |  | Flooding |  |  |  |
|  |  |  | Use concealment |  |  | Muzzle social media as a political force |  |  |  |
|  |  |  |  |  | Memes | Tertiary sites amplify news |  |  |  |
|  |  |  |  |  | Trial content | Twitter bots amplify |  |  |  |
|  |  |  |  |  |  | Twitter trolls amplify and manipulate |  |  |  |
|  |  |  |  |  |  | Use hashtag |  |  |  |

name
“murderers, thieves, and executioners”

description
A tweet published to the Twitter account of Iwona Michałek, Poland's deputy minister of development, labor, and technology, disseminated the false narrative that she no longer wanted to be affiliated with the PiS party. The tweet condemned PiS as the party of “murderers, thieves, and executioners” and featured a cartoon of PiS leader Jarosław Kaczyński in prison

Metadata

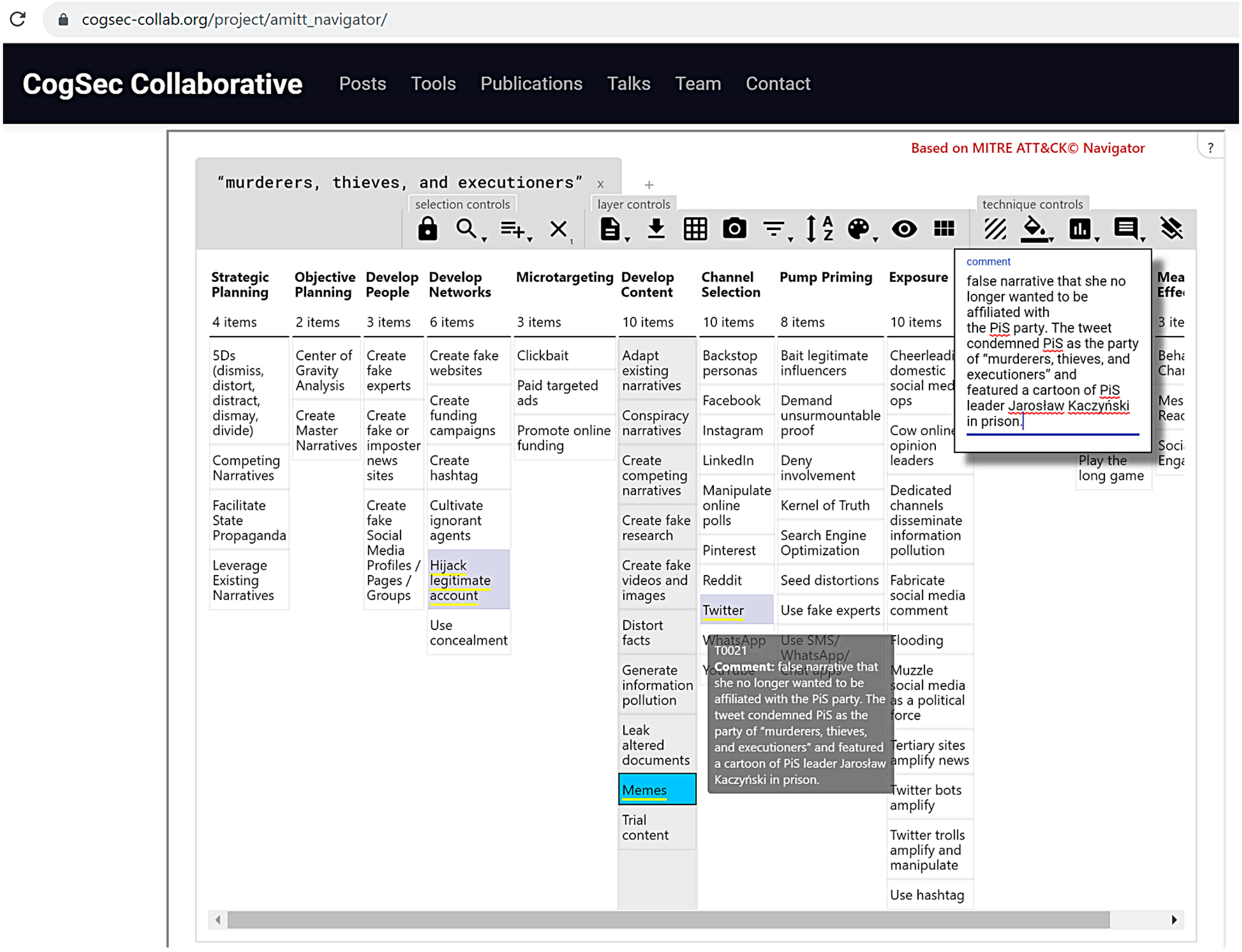
name
date

value
12 January 2021

remove this metadata

add more metadata

Figure 8. Quickly and easily add comments and customize



Based on MITRE ATT&CK© Navigator

“murderers, thieves, and executioners”

| Strategic Planning | Objective Planning | Develop People | Develop Networks | Microtargeting | Develop Content | Channel Selection | Pump Priming | Exposure |  | Mea Effe |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 items | 2 items | 3 items | 6 items | 3 items | 10 items | 10 items | 8 items | 10 items |  | 3 ite |
| 5Ds (dismiss, distort, distract, dismay, divide) | Center of Gravity Analysis | Create fake experts | Create fake websites | Clickbait | Adapt existing narratives | Backstop personas | Bait legitimate influencers | Cheerleadi domestic social med ops |  | Beha Char |
| Competing Narratives | Create Master Narratives | Create fake or imposter news sites | Create funding campaigns | Paid targeted ads | Conspiracy narratives | Facebook | Demand unsurmountable proof | Cow online opinion leaders |  | Mes Reac |
| Facilitate State Propaganda |  | Create fake Social Media Profiles / Pages / Groups | Create hashtag | Promote online funding | Create competing narratives | Instagram | Deny involvement | Dedicated channels disseminate information pollution | Play the long game | Soci Enga |
| Leverage Existing Narratives |  |  | Cultivate ignorant agents |  | Create fake research | LinkedIn | Kernel of Truth | Fabricate social media comment |  |  |
|  |  |  | Hijack legitimate account |  | Create fake videos and images | Manipulate online polls | Search Engine Optimization | Flooding |  |  |
|  |  |  | Use concealment |  | Distort facts | Pinterest | Seed distortions | Muzzle social media as a political force |  |  |
|  |  |  |  |  | Generate information pollution | Reddit | Use fake experts | Tertiary sites amplify news |  |  |
|  |  |  |  |  | Leak altered documents | Twitter | Use SMS/ WhatsApp/ Chat apps | Twitter bots amplify |  |  |
|  |  |  |  |  | Memes | WhatsApp |  | Twitter trolls amplify and manipulate |  |  |
|  |  |  |  |  | Trial content | YouTube |  | Use hashtag |  |  |

comment

false narrative that she no longer wanted to be affiliated with the PiS party. The tweet condemned PiS as the party of “murderers, thieves, and executioners” and featured a cartoon of PiS leader Jarosław Kaczyński in prison.

T0021

**Comment:** false narrative that she no longer wanted to be affiliated with the PiS party. The tweet condemned PiS as the party of “murderers, thieves, and executioners” and featured a cartoon of PiS leader Jarosław Kaczyński in prison.

As shown in Figure 7, comments that are added to individual techniques, using the visual interface, are included in the layer export (JSON and Excel). This makes it easier and faster for practitioners to produce data files that are codified to universal standards (e.g., STIX), as shown in Figure 8.

*Response*

The purpose of this scenario was to evaluate whether DISARM made it easy for practitioners to respond to disinformation attacks during the 'response' phase of disinformation defence. All objects in the DISARM framework are cross-referenced to parent and child relationships, offering a 'one-click' solution to counter-response planning. These relationships are hyperlinked in both the navigator web application and the GitHub repository. Table 7 shows the technique identified in the sighting phase (column one) and the corresponding screenshot from the GitHub repository shows the parent/child relationships and links (column 2).

*Analysis*

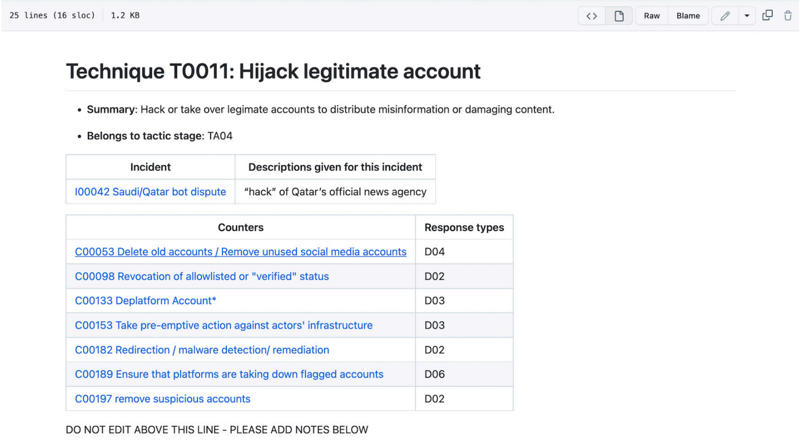
The purpose of this scenario was to evaluate whether DISARM made it easy for practitioners to share, record, and interrogate incident and/or campaign activity and measure impact during the 'analysis' phase of disinformation defence.

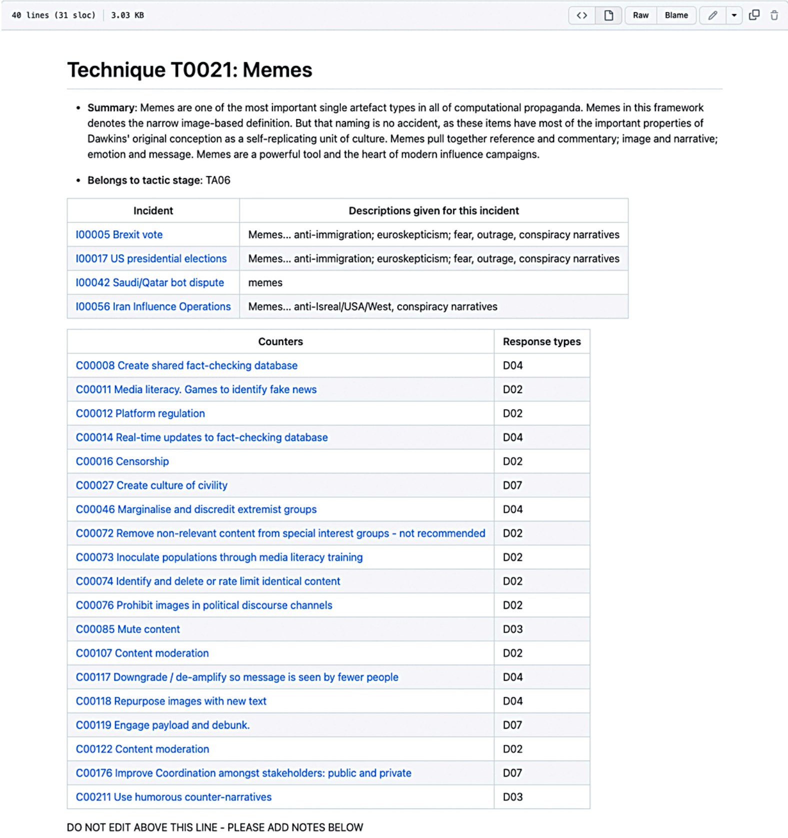
Currently, the DISARM framework has one tactic and three techniques defined for measuring effectiveness (TA12: Measure Effectiveness from the DISARM Red Framework). TA12 is about how the FIMI actors measure the effectiveness of their [disinformation] campaigns. However, TA12 could be used by FIMI threat analysts to measure the effectiveness of campaigns from the threat actors they are tracking as well as the effectiveness of the countermeasures the defenders implement. Similarly, strategic communications practitioners would find utility in TA12 due to the familiarity of these techniques (disinformation famously leverages the same microtargeting, tracking, and other platform affordances that make digital marketing so effective). Table 8 shows the tactic (column one), techniques (column two), and summary description of the technique (column three). The techniques cover three core components of impact measurement that marketers would be familiar with: T0062 Behaviour changes, T0063 Message reach, and T0064 Social media engagement.

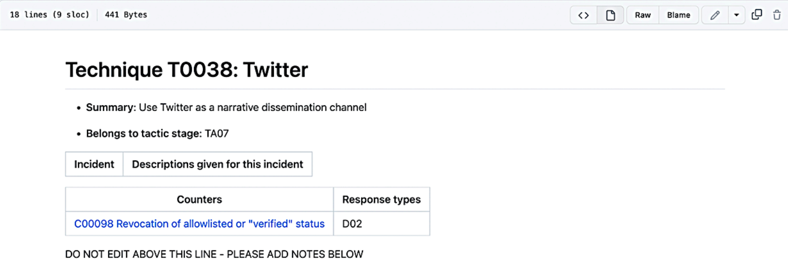
The natural step after measuring effectiveness is to benchmark its impact. Not only does this provide context and make measurements actionable, but it also prioritizes resourcing, developing the discipline, and informing investment decisions. Insofar as it is a framework, DISARM should not be expected to perform benchmarking. However, it could link to and/or be integrated with an existing mechanism, for example the MITRE ATTACK© Navigator web application and MISP intelligence platform have both been adapted for disinformation defence and link to DISARM documentation and resources.

During this phase of the assessment, no links were found from the DISARM documentation or resources to a universal benchmarking mechanism. Where it was developed and housed on

## Table 7. Response scenario evaluation process

<table>
<tr><th>Technique</th><th>Potential counters</th></tr>
<tr><td>Technique T0011: Hijack legitimate account</td><td>


<h3>Technique T0011: Hijack legitimate account</h3>
<ul>
<li><b>Summary:</b> Hack or take over legimate accounts to distribute misinformation or damaging content.</li>
<li><b>Belongs to tactic stage:</b> TA04</li>
</ul>
<table>
<tr><th>Incident</th><th>Descriptions given for this incident</th></tr>
<tr><td>I00042 Saudi/Qatar bot dispute</td><td>"hack" of Qatar's official news agency</td></tr>
</table>
<table>
<tr><th>Counters</th><th>Response types</th></tr>
<tr><td>C00053 Delete old accounts / Remove unused social media accounts</td><td>D04</td></tr>
<tr><td>C00098 Revocation of allowlisted or "verified" status</td><td>D02</td></tr>
<tr><td>C00133 Deplatform Account*</td><td>D03</td></tr>
<tr><td>C00153 Take pre-emptive action against actors' infrastructure</td><td>D03</td></tr>
<tr><td>C00182 Redirection / malware detection/ remediation</td><td>D02</td></tr>
<tr><td>C00189 Ensure that platforms are taking down flagged accounts</td><td>D06</td></tr>
<tr><td>C00197 remove suspicious accounts</td><td>D02</td></tr>
</table>
DO NOT EDIT ABOVE THIS LINE - PLEASE ADD NOTES BELOW
</td></tr>
<tr><td>Technique T0021: Memes</td><td>


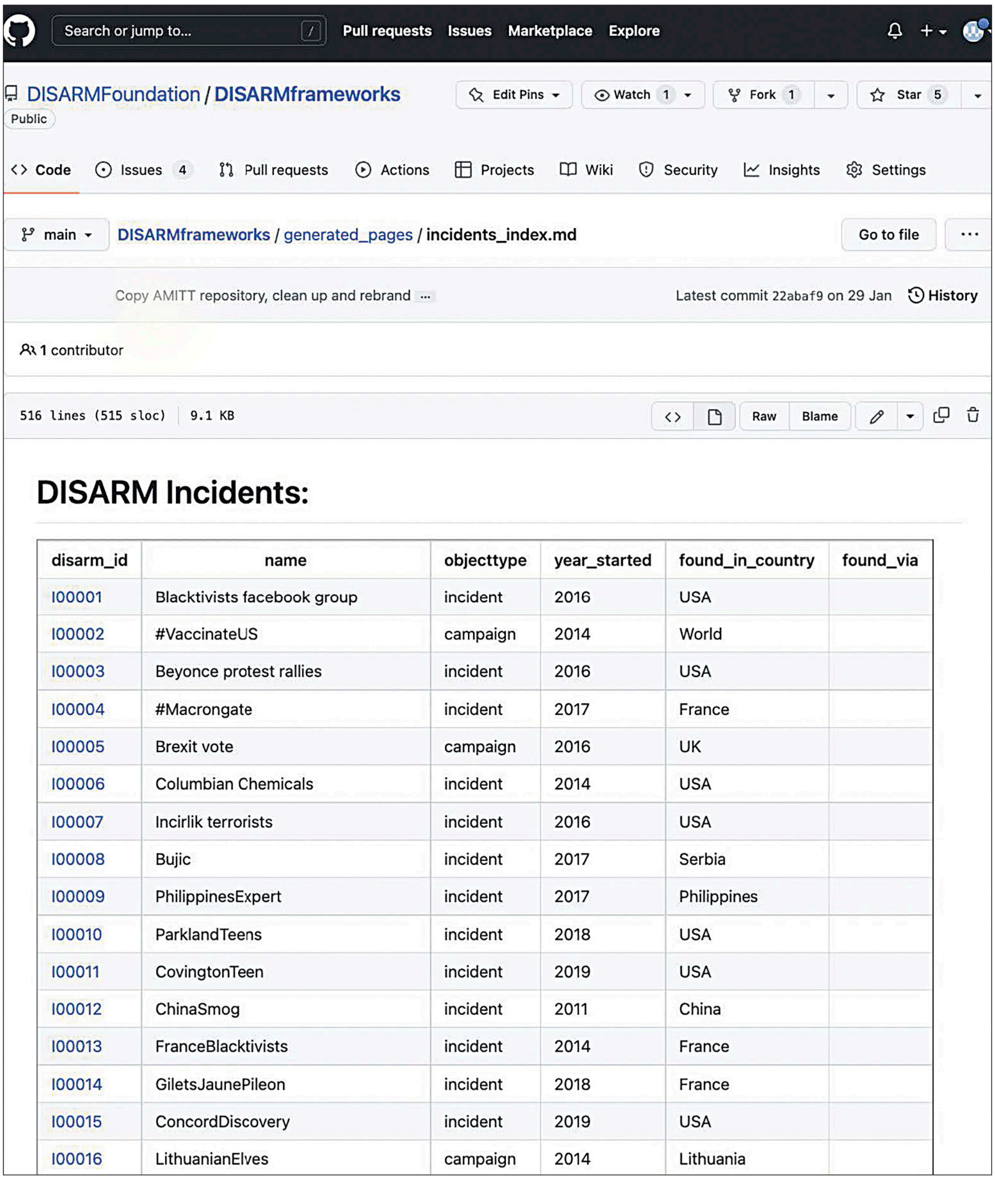
<h3>Technique T0021: Memes</h3>
<ul>
<li><b>Summary:</b> Memes are one of the most important single artefact types in all of computational propaganda. Memes in this framework denotes the narrow image-based definition. But that naming is no accident, as these items have most of the important properties of Dawkins' original conception as a self-replicating unit of culture. Memes pull together reference and commentary; image and narrative; emotion and message. Memes are a powerful tool and the heart of modern influence campaigns.</li>
<li><b>Belongs to tactic stage:</b> TA06</li>
</ul>
<table>
<tr><th>Incident</th><th>Descriptions given for this incident</th></tr>
<tr><td>I00005 Brexit vote</td><td>Memes... anti-immigration; euroskepticism; fear, outrage, conspiracy narratives</td></tr>
<tr><td>I00017 US presidential elections</td><td>Memes... anti-immigration; euroskepticism; fear, outrage, conspiracy narratives</td></tr>
<tr><td>I00042 Saudi/Qatar bot dispute</td><td>memes</td></tr>
<tr><td>I00056 Iran Influence Operations</td><td>Memes... anti-Isreal/USA/West, conspiracy narratives</td></tr>
</table>
<table>
<tr><th>Counters</th><th>Response types</th></tr>
<tr><td>C00008 Create shared fact-checking database</td><td>D04</td></tr>
<tr><td>C00011 Media literacy. Games to identify fake news</td><td>D02</td></tr>
<tr><td>C00012 Platform regulation</td><td>D02</td></tr>
<tr><td>C00014 Real-time updates to fact-checking database</td><td>D04</td></tr>
<tr><td>C00016 Censorship</td><td>D02</td></tr>
<tr><td>C00027 Create culture of civility</td><td>D07</td></tr>
<tr><td>C00046 Marginalise and discredit extremist groups</td><td>D04</td></tr>
<tr><td>C00072 Remove non-relevant content from special interest groups - not recommended</td><td>D02</td></tr>
<tr><td>C00073 Inoculate populations through media literacy training</td><td>D02</td></tr>
<tr><td>C00074 Identify and delete or rate limit identical content</td><td>D02</td></tr>
<tr><td>C00076 Prohibit images in political discourse channels</td><td>D02</td></tr>
<tr><td>C00085 Mute content</td><td>D03</td></tr>
<tr><td>C00107 Content moderation</td><td>D02</td></tr>
<tr><td>C00117 Downgrade / de-amplify so message is seen by fewer people</td><td>D04</td></tr>
<tr><td>C00118 Repurpose images with new text</td><td>D04</td></tr>
<tr><td>C00119 Engage payload and debunk.</td><td>D07</td></tr>
<tr><td>C00122 Content moderation</td><td>D02</td></tr>
<tr><td>C00176 Improve Coordination amongst stakeholders: public and private</td><td>D07</td></tr>
<tr><td>C00211 Use humorous counter-narratives</td><td>D03</td></tr>
</table>
DO NOT EDIT ABOVE THIS LINE - PLEASE ADD NOTES BELOW
</td></tr>
<tr><td>Technique T0038: Twitter</td><td>


<h3>Technique T0038: Twitter</h3>
<ul>
<li><b>Summary:</b> Use Twitter as a narrative dissemination channel</li>
<li><b>Belongs to tactic stage:</b> TA07</li>
</ul>
<table>
<tr><th>Incident</th><th>Descriptions given for this incident</th></tr>
</table>
<table>
<tr><th>Counters</th><th>Response types</th></tr>
<tr><td>C00098 Revocation of allowlisted or "verified" status</td><td>D02</td></tr>
</table>
DO NOT EDIT ABOVE THIS LINE - PLEASE ADD NOTES BELOW
</td></tr>
</table>

## Table 8. Analysis scenario evaluation process – DISARM techniques for measuring effectiveness

| Tactic | Technique | Summary |
|---|---|---|
| TA12: Measure Effectiveness | T0062: Behaviour changes | Monitor and evaluate behaviour changes from incidents |
| | T0063: Message reach | Monitor and evaluate message reach in incidents |
| | T0064: Social media engagement | Monitor and evaluate social media engagement in incident |

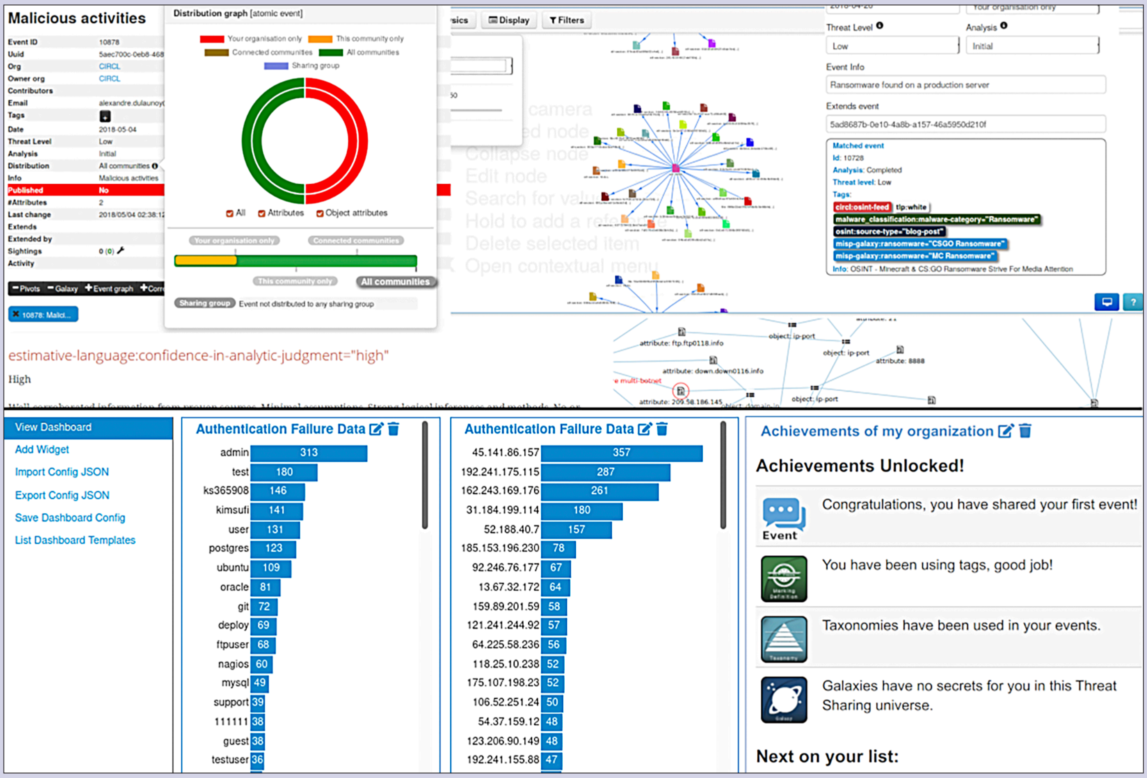
## Figure 9. Extract from list of recorded incidents on GitHub



### DISARM Incidents:

| disarm_id | name | objecttype | year_started | found_in_country | found_via |
|---|---|---|---|---|---|
| I00001 | Blacktivists facebook group | incident | 2016 | USA | |
| I00002 | #VaccinateUS | campaign | 2014 | World | |
| I00003 | Beyonce protest rallies | incident | 2016 | USA | |
| I00004 | #Macrongate | incident | 2017 | France | |
| I00005 | Brexit vote | campaign | 2016 | UK | |
| I00006 | Columbian Chemicals | incident | 2014 | USA | |
| I00007 | Incirlik terrorists | incident | 2016 | USA | |
| I00008 | Bujic | incident | 2017 | Serbia | |
| I00009 | PhilippinesExpert | incident | 2017 | Philippines | |
| I00010 | ParklandTeens | incident | 2018 | USA | |
| I00011 | CovingtonTeen | incident | 2019 | USA | |
| I00012 | ChinaSmog | incident | 2011 | China | |
| I00013 | FranceBlacktivists | incident | 2014 | France | |
| I00014 | GiletsJaunePileon | incident | 2018 | France | |
| I00015 | ConcordDiscovery | incident | 2019 | USA | |
| I00016 | LithuanianElves | campaign | 2014 | Lithuania | |

Figure 10. Screenshot of MISP data visualization



Malicious activities

| | |
|---|---|
| Event ID | 10878 |
| Uuid | 5aec700c-0eb8-468 |
| Org | CIRCL |
| Owner org | CIRCL |
| Contributors | |
| Email | alexandre.dulaunoy |
| Tags | |
| Date | 2018-05-04 |
| Threat Level | Low |
| Analysis | Initial |
| Distribution | All communities |
| Info | Malicious activities |
| Published | No |
| #Attributes | 2 |
| Last change | 2018/05/04 02:38:1 |
| Extends | |
| Extended by | |
| Sightings | 0 (0) |
| Activity | |

Sharing group: Event not distributed to any sharing group

estimative-language:confidence-in-analytic-judgment="high"

High

Threat Level: Low

Analysis: Initial

Event Info: Ransomware found on a production server

Extends event: 5ad8687b-0e10-4a8b-a157-46a5950d210f

Matched event

Id: 10728

Analysis: Completed

Threat level: Low

Tags: circl:osint-feed tlp:white malware_classification:malware-category="Ransomware" osint:source-type="blog-post" misp-galaxy:ransomware="CSGO Ransomware" misp-galaxy:ransomware="MC Ransomware"

Info: OSINT - Minecraft & CS:GO Ransomware Strive For Media Attention

Achievements of my organization

Achievements Unlocked!

Congratulations, you have shared your first event!

You have been using tags, good job!

Taxonomies have been used in your events.

Galaxies have no secrets for you in this Threat Sharing universe.

Next on your list:

GitHub, DISARM is open source and can, in the absence of third-party tools, be used to build a knowledge base of recorded disinformation incidents and campaigns, as shown in Figure 9.

The DISARM framework is not capable of storing and visualizing data. It was designed to work in harmony with existing information security frameworks and languages and would be highly familiar to practitioners just as it would enable third-party developers to adapt their current solutions to accommodate DISARM (e.g., MITRE ATT&CK© Navigator and MISP).

To be clear, MISP is an open-source threat intelligence and platform for sharing cyber security-related information and has been used to share disinformation data. One of many such platforms, MISP allows practitioners to share, store, correlate, and analyze targeted attacks. It also includes data visualization options and a customizable data dashboard, as shown in Figure 10.

## Validation

In this phase, DISARM's usability was assessed against recognized usability principles (i.e., 'Usability Heuristics'). Test criteria are presented in Table 9.

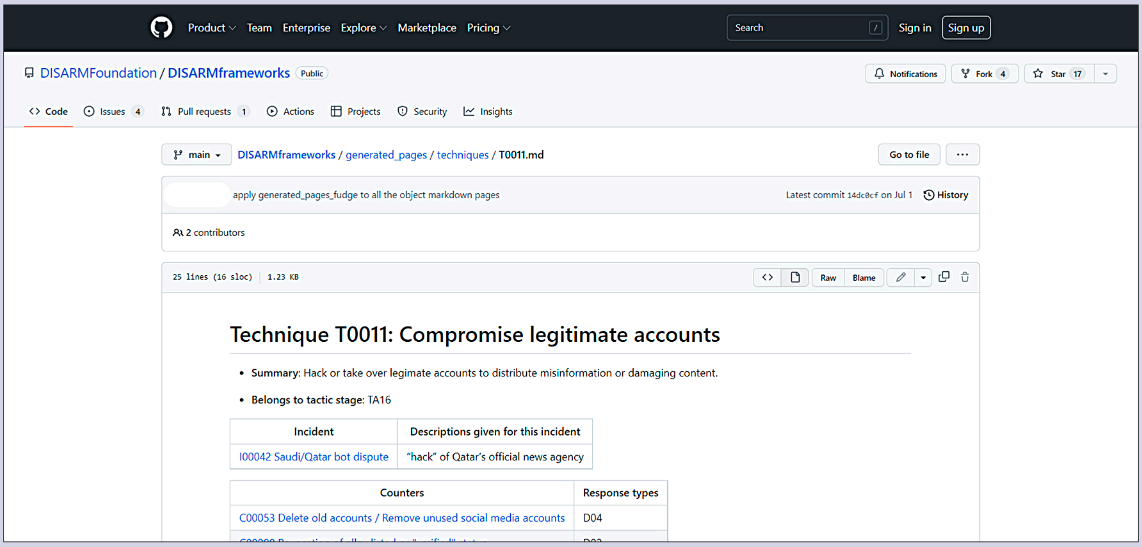
### *Visibility of system status*

DISARM was developed on GitHub, an open-source community of over 40 million developers. It was designed to enable collaborative coding with team spaces and audit trails by named users. Information wayfinding (e.g., breadcrumbs, intuitive navigation, icons, and searches) and system status information (e.g., contributors, last comment, and history) keep users informed about what is going on through appropriate feedback within a reasonable amount of time. For reference, the user's view of DISARM on GitHub is presented in Figure 11.

### Table 9. 'Usability Heuristics' criteria

| | Title | Criteria |
|---|---|---|
| 1. | **Visibility of system status** | The design should keep users informed about what is going on, through appropriate feedback and within a reasonable amount of time. |
| 2. | **Match between system and the real world** | The design should speak the user's language (i.e., use words, phrases, and concepts that are familiar to the user, rather than jargon) and follow real-world conventions, making information appear in a narrative form and logical order. |
| 3. | **User control and freedom** | Since users often perform actions by mistake, they need a clearly marked "emergency exit" to leave the unwanted action without having to go through an extended process. |
| 4. | **Consistency and standards** | Users should not have to wonder whether different words, situations, or actions mean the same thing (per platform and industry conventions). |
| 5. | **Error prevention** | While good error messages are important, the best designs prevent problems from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action. |
| 6. | **Recognition rather than recall** | Minimize the user's cognitive load by making elements, actions, and options visible. The user should not have to remember information across interfaces. Information required to use the design (e.g., field labels or menu items) should be visible or easily retrievable when needed. |
| 7. | **Flexibility and efficiency of use** | Shortcuts – hidden from novice users – may speed up interactions for expert users such that the design can cater to both inexperienced and experienced users. Allow users to tailor frequent actions to their needs. |
| 8. | **Aesthetic and minimalist design** | Interfaces should not contain information which is irrelevant or rarely needed. Every extra unit of information in an interface competes with the relevant units of information and diminishes visibility. |
| 9. | **Help users recognize, diagnose, and recover from errors** | Error messages should be expressed in plain language (no error codes), clearly note the problem, and offer a constructive solution. |
| 10. | **Help and documentation** | It is best if the system does not require any additional explanation. However, it may be necessary to provide documentation to help users understand how to complete their tasks. |

Figure 11. Screenshot of DISARM on GitHub



apply generated_pages_fudge to all the object markdown pages

Latest commit 14dc0cf on Jul 1

2 contributors

25 lines (16 sloc) | 1.23 KB

### Technique T0011: Compromise legitimate accounts

- **Summary:** Hack or take over legimate accounts to distribute misinformation or damaging content.
- **Belongs to tactic stage:** TA16

| Incident | Descriptions given for this incident |
|---|---|
| I00042 Saudi/Qatar bot dispute | "hack" of Qatar's official news agency |

| Counters | Response types |
|---|---|
| C00053 Delete old accounts / Remove unused social media accounts | D04 |

*Match between system and the real world*
By combining standard defence and communications terminology, DISARM is a framework that is both intuitive and familiar to practitioners. For example, Table 10 compares the similarities by communications phase (column one) between the terminology used by the US Department of Defense (column two), the Chartered Institute of Marketing (column three), and DISARM (column four).

Table 11 shows, by stage, the similarities between the Disinformation Kill Chain and a selection of communications stimulus-response models that demonstrate the attitudinal and behavioural changes in response to persuasive communication: the attention, interest, desire and action (AIDA); the Hierarchy of Effects model of cognitive, affective, and conative stages; the Consumer Adoption process of awareness, interest, evaluation, trial, and adoption; and the Information Process that includes Exposure, Attention, Comprehension, Acceptance and Retention traits.

The action steps of the Disinformation Kill Chain are included in Table 11 to enable like-for-like comparison. Figure 13 presents the process.

*User control and freedom*
There are three methods for interfacing with the DISARM framework: raw code, documents, and navigator web application.

- *Raw code*: GitHub offers a suite of user tools for coding, including a range of editing, undo, and redo options that follow best practices.
- *Documents*: the repository provides access to all DISARM objects that are hyperlinked and cross-referenced. To enable rollback, the site contains a full audit trail of changes.
- *Navigator web application*: has been adapted for DISARM, and there is no 'undo' button.

Table 10. Comparing similarities in terminology by communications phase

| Communication phases | US Department of Defense | Chartered Institute of Marketing[58] | DISARM[59] |
|---|---|---|---|
| PLAN[60]<br>Envision the desired outcome. Lay out effective ways of achieving it. Communicate the vision, intent, and decisions, focusing on expected results. | TACTICS[61]<br>Strategic use and ordered arrangement of forces in relation to each other. | MARKETING STRATEGY<br>The set of objectives that an organization allocates to its marketing arm to support the overall corporate strategy, together with the broad methods chosen to achieve these objectives. | TACTICS<br>Activities that someone disseminating is likely to perform. |
| PREPARE[62]<br>Activities conducted pre-execution to improve outcomes (e.g., develop the ecosystem needed to support the action: people, network, channels, content). | TECHNIQUES[63]<br>Non-prescriptive methods used to perform missions, functions, or tasks. | MARKETING PLAN<br>A written plan that describes all activities involved in achieving a particular marketing objective, and their relationship to one another in both time and importance. | TECHNIQUES<br>Activities that might be seen at each phase. |
| EXECUTE[64]<br>Run the action, from initial exposure to wrap-up and/or maintenance. | PROCEDURES[65]<br>Standard, detailed steps that prescribe how to perform specific tasks. | IMPLEMENTATION<br>Delivery phase of various marketing models. | TASKS<br>Things that need to be done at each phase. |
| ASSESS[66]<br>Evaluate effectiveness of action for future planning purposes. | Not Available / Applicable | MARKETING METRICS<br>Measurements that help to quantify marketing performance. | MEASURE EFFECTIVENESS<br>Measure effectiveness of incident for future planning purposes. |

*Consistency and standards*

The DISARM framework was built by the information security community and is based on pre-existing standards, models, and platforms, including MITRE ATT&CK©, STIX, GitHub, marketing models (Table 10), and terminology (Table 11).

*Error prevention*

Plain language is used throughout DISARM's documentation and guidelines to reinforce learning (e.g., "tasks are things you do, techniques are how you do them"[67]). Within the navigator web application, an error message is displayed if a user attempts to leave the page without saving.

[58] See https://marketingexpert.cim.co.uk/glossary/.

[59] See https://github.com/DISARMFoundation/DISARMframeworks.

[60] See https://github.com/DISARMFoundation/DISARMframeworks/blob/main/generated_pages/phases/P01.md.

[61] Headquarters Department of the Army, Army Doctrine Publication No. 1–01, Washington, DC: Army Publishing Directorate, 2019.

[62] See https://github.com/DISARMFoundation/DISARMframeworks/blob/main/generated_pages/phases/P02.md.

[63] Headquarters Department of the Army, Army Doctrine Publication No. 1–01.

[64] See https://github.com/DISARMFoundation/DISARMframeworks/blob/main/generated_pages/phases/P03.md.

[65] Headquarters Department of the Army, Army Doctrine Publication No. 1–01.

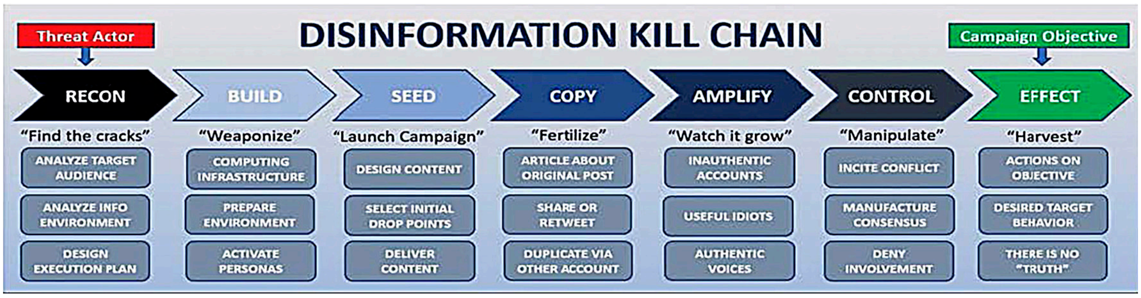
[66] See https://github.com/DISARMFoundation/DISARMframeworks/blob/main/generated_pages/phases/P04.md.

[67] See https://github.com/DISARMFoundation/DISARMframeworks.

Table 11. Comparison of similarities in disinformation and communications models

| Stage | Disinformation Kill Chain[68] | AIDA[69] | Hierarchy of Effects[70] | Consumer Adoption[71] | Information Process[72] |
|---|---|---|---|---|---|
| **Cognitive (logic/ think)** | Seed<br>↓<br>Copy | Attention | Awareness<br>↓<br>Knowledge | Awareness | Presentation<br>↓<br>Attention<br>↓<br>Comprehension |
| **Affective (emotions/ feel)** | Amplify | Interest<br>↓<br>Desire | Liking<br>↓<br>Preference<br>↓<br>Conviction | Interest<br>↓<br>Evaluation | Yielding<br>↓<br>Retention |
| **Conative (behaviour/ do)** | Control<br>↓<br>Effect | Action | Purchase | Trail<br>↓<br>Adoption | Behaviour |

Figure 12. Planning and action process of Disinformation Kill Chain



Source: Mitre Corporation

[68] U.S. Department of Homeland Security, *Combatting Targeted Disinformation Campaigns* (Washington, DC: U.S. Government Publishing Office, 2019), p. 16, https://www.dhs.gov/sites/default/files/publications/ia/ia_combatting-targeted-disinformation-campaigns.pdf.

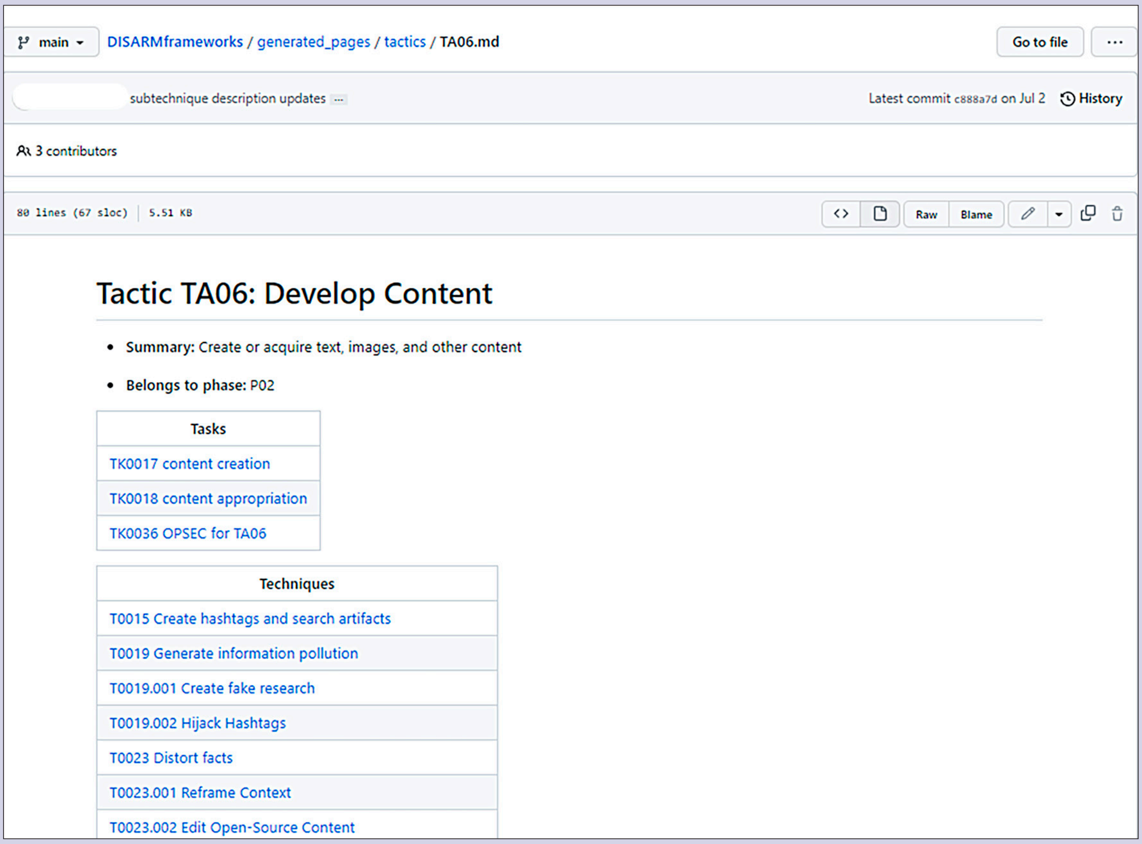
[69] Kobby Mensah and Fortune Amenuvor, 'The influence of marketing communications strategy on consumer purchasing behaviour in the financial services industry in an emerging economy', *Journal of Financial Services Marketing*, Volume 27 (2022):190–205, https://doi.org/10.1057/s41264-021-00121-0.

[70] Robert Lavidge and Gary Steiner, 'A Model for Predictive Measures of Advertising Effectiveness', *Journal of Marketing* (1961): 59–62.

[71] Philip Kotler et al., *Principles of Marketing*, Second European Edition (Prentice Hall Europe, 1999), p. 260.

[72] William McGuire, 'An Information Processing Model of Advertising Effectiveness', in *Behavioral and Management Science in Marketing* (New York: John Wiley, 1978), 156–80.

## Figure 13. Parent and child referencing


main
DISARMframeworks / generated_pages / tactics / TA06.md
Go to file
···
subtechnique description updates ···
Latest commit c888a7d on Jul 2
History
3 contributors
80 lines (67 sloc) | 5.51 KB
Raw
Blame

# Tactic TA06: Develop Content

- **Summary:** Create or acquire text, images, and other content
- **Belongs to phase:** P02

| Tasks |
|---|
| TK0017 content creation |
| TK0018 content appropriation |
| TK0036 OPSEC for TA06 |

| Techniques |
|---|
| T0015 Create hashtags and search artifacts |
| T0019 Generate information pollution |
| T0019.001 Create fake research |
| T0019.002 Hijack Hashtags |
| T0023 Distort facts |
| T0023.001 Reframe Context |
| T0023.002 Edit Open-Source Content |

*Recognition rather than recall*
Within the GitHub repository and the navigator web application, the information required to use the design is visible (when needed) and easily retrievable. The screenshots (Figures 13 and 14) show how each object contains parent and child information to make it easy to navigate the interface, without having to rely on memory.

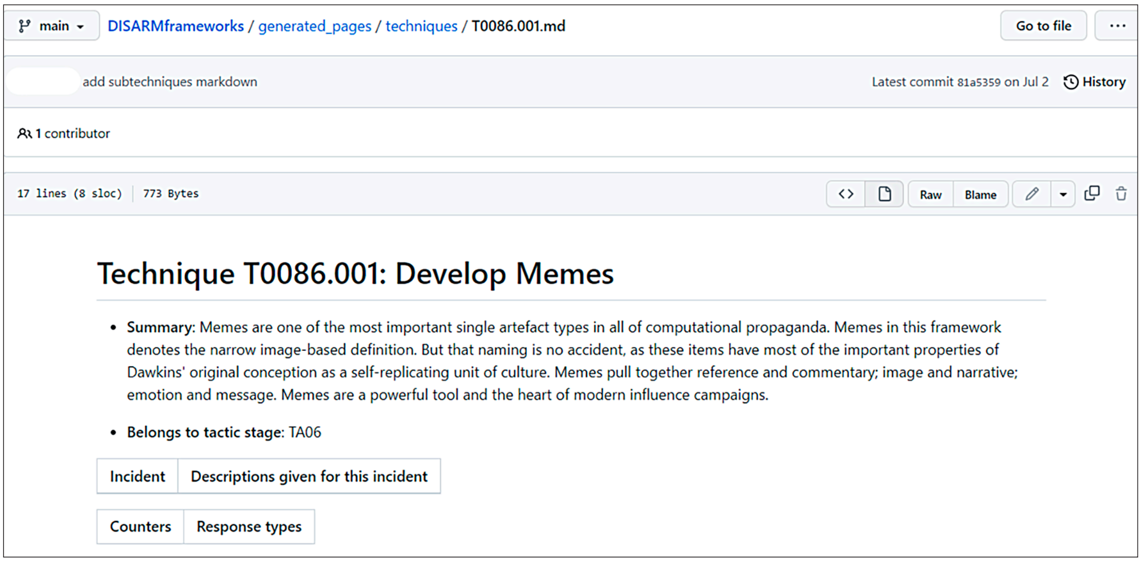
*Flexibility and efficiency of use*
GitHub's desktop and mobile version contains shortcuts and custom actions. The navigator web application allows for customization of scoring, colours, metadata fields, multiple import/export options, and more technical (nested layers) for advanced users. Although there are no macro recording features to tailor frequent actions, it is possible to create custom layers and templates for future use, as shown in Figure 15.

*Aesthetic and minimalist design*
The DISARM framework, GitHub repository, and navigator web application employ user interfaces that are designed to provide the most relevant and useful links, information, wayfinding, and language. Ample white space increases visibility between elements.

*Help users recognize, diagnose, and recover from errors*
This heuristic is not applicable to the evaluation because DISARM is not a software application.

*Help and documentation*
Help is available within the navigator web app, as shown in Figure 16.

Figure 14. Screenshot of parent and child referencing


main
DISARMframeworks / generated_pages / techniques / T0086.001.md
Go to file
add subtechniques markdown
Latest commit 81a5359 on Jul 2
History
1 contributor
17 lines (8 sloc) | 773 Bytes
Raw
Blame

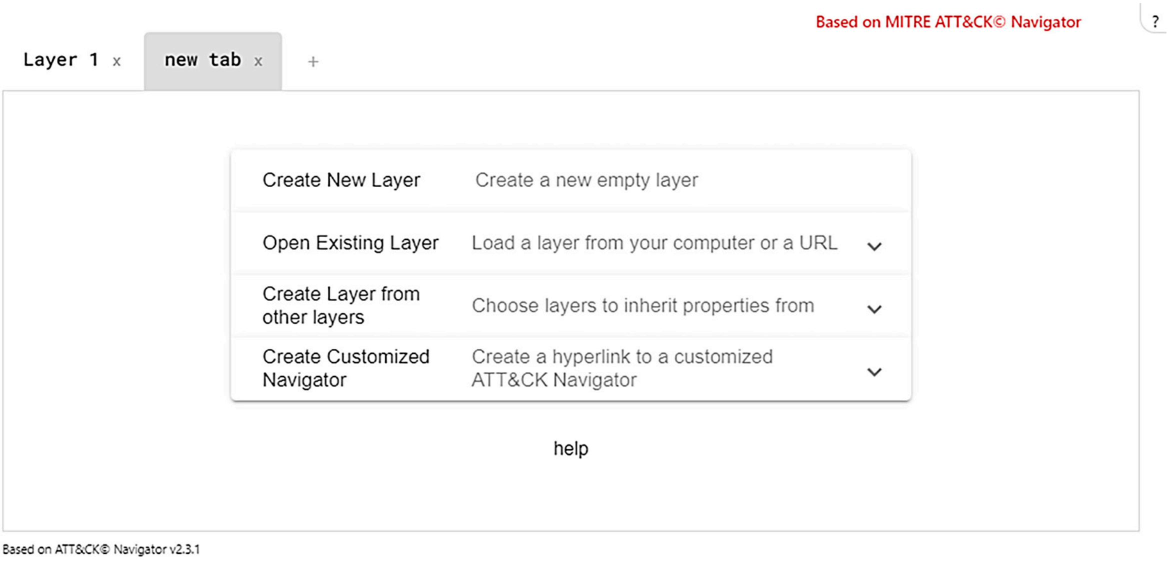
## Technique T0086.001: Develop Memes

- **Summary**: Memes are one of the most important single artefact types in all of computational propaganda. Memes in this framework denotes the narrow image-based definition. But that naming is no accident, as these items have most of the important properties of Dawkins' original conception as a self-replicating unit of culture. Memes pull together reference and commentary; image and narrative; emotion and message. Memes are a powerful tool and the heart of modern influence campaigns.
- **Belongs to tactic stage**: TA06

| Incident | Descriptions given for this incident |
|---|---|

| Counters | Response types |
|---|---|

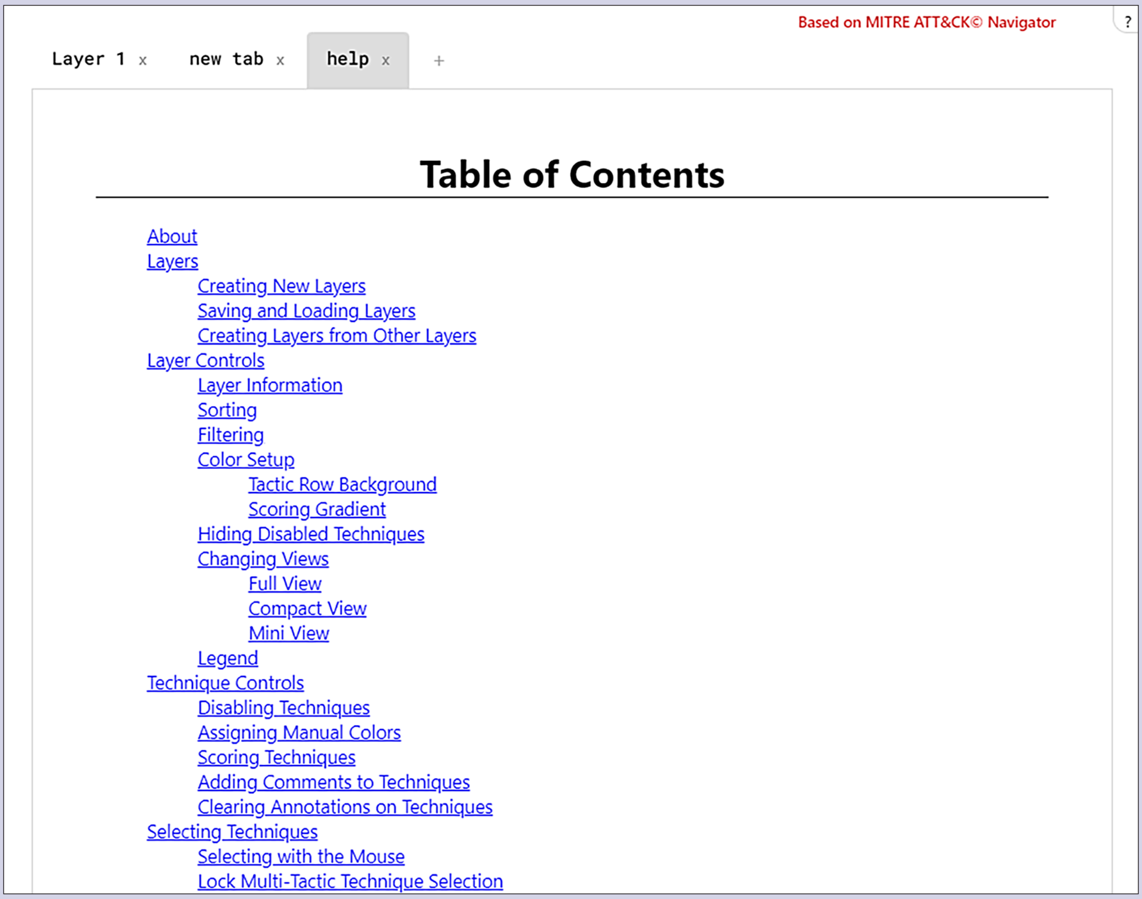
Figure 15. Screenshot navigator web application custom layers


Based on MITRE ATT&CK© Navigator
?
Layer 1 x
new tab x
+

| | |
|---|---|
| Create New Layer | Create a new empty layer |
| Open Existing Layer | Load a layer from your computer or a URL |
| Create Layer from other layers | Choose layers to inherit properties from |
| Create Customized Navigator | Create a hyperlink to a customized ATT&CK Navigator |

help
Based on ATT&CK© Navigator v2.3.1

Figure 16. Screenshot of help in navigator web app



# Table of Contents

- About
- Layers
  - Creating New Layers
  - Saving and Loading Layers
  - Creating Layers from Other Layers
- Layer Controls
  - Layer Information
  - Sorting
  - Filtering
  - Color Setup
    - Tactic Row Background
    - Scoring Gradient
  - Hiding Disabled Techniques
  - Changing Views
    - Full View
    - Compact View
    - Mini View
  - Legend
- Technique Controls
  - Disabling Techniques
  - Assigning Manual Colors
  - Scoring Techniques
  - Adding Comments to Techniques
  - Clearing Annotations on Techniques
- Selecting Techniques
  - Selecting with the Mouse
  - Lock Multi-Tactic Technique Selection

The community has created various playbooks, which are available from multiple online sources (e.g., the GitHub repository). They include introductions to the origin, content, and use of associated DISARM disinformation models: STIX, TTPs, and Countermeasures. This list of documents is shown below in Table 12.

A simulator was created using two frameworks to accelerate learning: red (offence, disinformation creator), and blue (defence, disinformation responder). Thus, practitioners who are new to DISARM and/or disinformation defence can practise and apply what they have learned in a realistic setting. The two frameworks can be seen in Figures 2 and 3.

By building a repository of campaigns, incidents, and related tactics, techniques and counters, GitHub's repository provides a knowledge base of real-world disinformation activity that can be used for learning and capability development.

### Credibility

In this evaluation, DISARM's credibility was assessed against recognized credibility principles (i.e., 'ISEAL Credibility Principles'). The principles were adapted from their original sustainability focus to be compatible with the DISARM assessment. Test criteria are presented in Table 13. The principle number appears in column one, the title in column two, and the criteria in column 3.

Table 12. List of guides and resources

| Title | Summary |
|---|---|
| DISARM Design Guide | Design and philosophy behind DISARM frameworks |
| DISARM User Guide | Ways to work with DISARM frameworks |
| DISARM TTP Guide | Describes each of the DISARM and counter TTPs |
| Proposed changes to DISARM | List of modifications (TTP, incident, and structure) being considered |
| DISARM Use Cases | Examples |
| DISARM Incident List | The incident descriptions we used to create DISARM |

Table 13. 'Credibility Principles' criteria (adapted)

| | Title | Criteria |
|---|---|---|
| 1. | **Impacts** | A credible system makes an impact where it matters. |
| 2. | **Collaboration** | A credible system works with others to create change. |
| 3. | **Value creation** | A credible system adds value. |
| 4. | **Measurable progress** | A credible system can demonstrate the difference it is making. |
| 5. | **Stakeholder engagement** | A credible system listens and learns. |
| 6. | **Transparency** | A credible system earns trust by being open and honest. |
| 7. | **Impartiality** | A credible system is impartial. |
| 8. | **Reliability** | A credible system provides trustworthy assessments of users' performance. |
| 9. | **Truthfulness** | A credible system's claims and communications can be trusted. |
| 10. | **Continuous improvement** | A credible system keeps improving. |

Table 14. Key findings of usability evaluation

| Test | Title | Result | Key findings |
|---|---|---|---|
| 1. | **Visibility of system status** | Present | GitHub developer platform provides full tracking. |
| 2. | **Match between system and real world** | Present | DISARM combines standard strategic communication and defence terminology. |
| 3. | **User control and freedom** | Present | 'Undo' is not applicable, but a clear hierarchical structure and cross-referencing are present. |
| 4. | **Consistency and standards** | Present | It follows communications industry conventions. |
| 5. | **Error prevention** | Present | Ordinary language was used to reinforce learning (e.g., 'tasks are things you do, techniques are how you do them'). |
| 6. | **Recognition rather than recall** | Present | Clear use of 'ID', 'name', and 'summary' present throughout process. |
| 7. | **Flexibility and efficiency of use** | Present | Provides option for use to access raw data file and HTML sheets. |
| 8. | **Aesthetic and minimalist design** | Present | Content and visual design are focused on essentials. |
| 9. | **Recognize, diagnose, and recover from errors** | N/A | No 'error messages' appear, as this is a methodology not a software. |
| 10. | **Help and documentation** | Present | User guide, instructions for updating major and minor changes, training tool (red/blue frameworks), spurs collaboration, full tracking of changes, and proposed iterations. |

## Results

### Key findings: pre-evaluation

**Imperative for universal approach:** In the wake of a rapidly evolving threat management ecosystem, DISARM was built by and for the security community.

**Robust and future-proof design:** DISARM is a living tool with a clear roadmap. Built on GitHub, it allows for a full audit history, collaborative development, and a repository for documentation and raw code. The structure was designed to enable scalability (as the discipline evolves) and integration with third-party developers to create software applications to record, process, and visualize information activity.

**ATTACK scenario tool (Figure 3) accelerates learning:** Capability development has been built alongside the framework and includes a web application to simulate war games (Red/Blue attack/counter), user guides, and independent playbooks by the security community.

### Key findings: usability evaluation

**Intuitive learning for ease of use:** There is an instant familiarity with DISARM, as the language and framework reflect the terminology and campaign structure of standard communications practice.

**Compatible with existing tools in the information security ecosystem:** Codifying the incident in the STIX language and serialization format was simplified by using the DISARM framework to identify the tactics and techniques found in the case study.

**Compliant with best practices for ease-of-use:** DISARM is compliant with the nine applicable heuristic tests (see Table 14). The test number is in column one, the title in column two, the result in column three, and the key findings in column four.

**Key findings: credibility evaluation**

**Credible system for a universal approach:** DISARM is compliant with all 10 credibility principles (see Table 15). The test number is in column one, the title in column two, the result  in column three, and the key findings in column four.

Table 15. Key findings credibility evaluation

| Test | Title | Result | Key findings |
|---|---|---|---|
| 1. | **Impacts** | Compliant | DISARM has a clear purpose ('to describe and understand disinformation incidents'). It defines and clearly communicates its scope, its specific objectives, and its strategies for achieving these objectives. The system focuses on the significant disinformation impacts in its scope. DISARM adopts international norms and can be adapted to local or sector-specific conditions where this helps improve impact. |
| 2. | **Collaboration** | Compliant | DISARM has been designed to serve a diverse stakeholder group, including the global counter-disinformation community. AMITT was originally developed in 2019 by the Credibility Coalition's Misinfosec Working Group (MisinfosecWG), with inputs from the misinfosec community. |
| 3. | **Value creation** | Compliant | The DISARM framework is open source and licensed under Creative Commons. It was designed to work with existing tools and is readily available on GitHub developer platform. DISARM's style is based on the MITRE ATT&CK framework; STIX templates for DISARM objects are available in the DISARM_CTI repo – these make it easy for DISARM data to be passed between Information Sharing and Analysis Organizations using standards like TAXII. |
| 4. | **Measurable progress** | Compliant | The uniform nature of DISARM enables consistent monitoring, managing, and reporting. It facilities the ability for third-party application developers to create data storage and visualization tools to record and share disinformation incidents in one universal language. |
| 5. | **Stakeholder engagement** | Compliant | DISARM was the result of a multi-stakeholder task force and continues as an open and inclusive group. GitHub is the primary platform for engagement. A Google form is also available, a new website and an updated version of the framework based on stakeholder feedback and research were launched during the writing of this report. |

| Test | Title | Result | Key findings |
|---|---|---|---|
| 6. | **Transparency** | Compliant | A full history of the origins of the DISARM system, minor and major changes, and the raw source code are available on GitHub. Information on who has responsibility for ongoing updates and changes is also available there. |
| 7. | **Impartiality** | Compliant | DISARM originated as and remains an open-source system with decentralized ownership. All changes are open, transparent and capture an audit trail of edits (via GitHub). |
| 8. | **Reliability** | Compliant | DISARM has clearly defined guides, frameworks, and objects (phases, tactics, techniques, tasks, counters, actors, response types, meta techniques, and incidents) to ensure that it is consistently implemented and assessed. |
| 9. | **Truthfulness** | Compliant | DISARM and its repository on GitHub do not make any unsubstantiated claims. All information is tagged with metadata and all documentation contains suitable referencing where applicable. |
| 10. | **Continuous improvement** | Compliant | DISARM has clearly defined ownership of ongoing maintenance and development. Full details are available, with contact information and useful links to relevant parties on GitHub. |

# 6. Recommendations

FIMI is a growing political and security challenge[73] and there is a need for a common defence framework. A whole-of-society approach to increase resilience against and to counter FIMI is required to leverage the different capabilities and competences that lie within governments, civil society, and private industry. In this evaluation of the rapid adoption of DISARM, consideration was also given to how wider adoption of the framework might be achieved, and four recommendations are offered:

**1. Apply the DISARM framework.**
The DISARM framework should be used by FIMI threat analysts and strategic communications practitioners within government, international organizations and institutions, platforms, academia, private industry, and civil society.

**2. Engage private industry in software development.**
Engage private industry in developing application software to record, process, and visualize information activity in accordance with the DISARM framework. Commercializing this emerging discipline and fostering innovation will further strengthen the collective defence against FIMI. Businesses have an invaluable role to play in supporting the wider adoption of DISARM by developing software applications that can make data usable and reveal its value, enabling stakeholders to record, process, and visualize FIMI activity. As online FIMI practices are built upon contemporary communications practices, there is an opportunity for developers to integrate DISARM components into existing automation solutions. This could fast-track DISARM's adoption amongst private industry practitioners, who are already familiar with the solutions, and thus popularize FIMI defence practices.

**3. Establish FIMI monitoring and analysis as a communications discipline.**
Establish consistent threat management work streams/specialties in accordance with the 'sighting', 'response', and 'analysis' process for the introduction of robust defence teams that are applicable across government as well as in private industry. It is envisaged that these organizations already have individuals and teams who can be upskilled to meet the rising demand for FIMI defence. Plotting a capability matrix and conducting a skills mapping exercise for the three key disciplines would be the initial steps to identify areas for development within the organizations.

**4. Prepare future generations of practitioners and the public.**
Partner with academia and professional bodies to deliver age-appropriate tuition. Just as social media managers and content developers are now mainstream disciplines within the field of communications, FIMI defence specialists will soon take their place amongst them. Partnering with professional bodies and academia to teach a universal approach could accelerate the mobilization of defence, increase the talent pool, and stimulate innovation in this emerging discipline.

[73] European External Action Service (EEAS), 'Tackling Disinformation'.

# 7. Conclusions

A practitioner with a fundamental skillset in strategic communications would find DISARM intuitive to learn and easy to use. A FIMI threat analyst can feel reassured that DISARM is a credible system for a universal approach to catalogue and plan disinformation threats and attacks. Furthermore, the capability development tools would facilitate mass adoption and capacity building that is beneficial to the framework and its stakeholders.

Since DISARM was built by experts across the international security community, the final output is well-considered, practical, and fit for purpose. The decision to develop the framework collaboratively and make it free on GitHub allows independent third-party businesses to develop software applications that can unlock the DISARM framework's greater potential.

Using an established heuristic model, this report has shown that a strategic communications practitioner who has no previous knowledge of DISARM, but who knows how to identify fundamental communications techniques, can readily adopt and apply DISARM in their daily operations. Since the framework was created within the security community, it is a robust and market-ready solution that is relevant to immediate threats and can scale as the discipline evolves and expands. The heuristic assessment revealed several significant contributing factors in ensuring DISARM's suitability for practical application and rapid adoption. This includes intuitive language and structure that is native to the communications profession, and the suite of complementary capability development tools (e.g., navigator web app, countermeasures scenario-building and practices, and adversarial playbooks). Furthermore, this report has shown that DISARM works in harmony with existing tools in the information security ecosystem. FIMI is increasingly seen within geopolitical strategies to achieve strategic gains. External efforts to manipulate public opinion aim to shape attitudes and behaviours, including the weakening of public trust in institutions. A collective whole-of-society approach and immediate action are paramount in mitigating these urgent real-world problems on a scale that can only be met by a united global response. To achieve this, practitioners within government, international organizations and institutions, platforms, academia, private industry, and civil society must be encouraged to adopt the DISARM framework, and to support its further development.

# Annex I: Usability Heuristics

**'Usability Heuristics'**
This model, originally developed in 1990 by Jakob Nielsen and Rolf Molich,[74] was later refined by Nielsen based on a factor analysis of 249 usability problems[75] that presents a set of heuristics with maximum explanatory power. While there has been a slight refinement of the language used in the definitions, the 10 heuristics have remained relevant and unchanged since 1994. The principle number appears in column one, the title in column two, the criteria in column three, and the impact in column four.

### Table 16. 'Usability Heuristics' criteria and impact

| | Title | Criteria | Impact |
|---|---|---|---|
| 1. | **Visibility of system status** | The design should always keep users informed about what is going on, through appropriate feedback within a reasonable amount of time. | *When users know the current system status, they learn the outcome of their prior interactions and determine next steps. Predictable interactions create trust in the product as well as the brand.* |
| 2. | **Match between system and the real world** | The design should speak the user's language. Use words, phrases, and concepts that are familiar to the user, rather than jargon. Follow real-world conventions, making information appear in a natural and logical order. | *The way you should design depends very much on your specific users. Terms, concepts, icons, and images that seem perfectly clear to you and your colleagues may be unfamiliar or confusing to your users.*<br>*When a design's controls follow real-world conventions and correspond to desired outcomes (called natural mapping), it's easier for users to learn and remember how the interface works. This helps to build an experience that feels intuitive.* |
| 3. | **User control and freedom** | Users often perform actions by mistake. They need a clearly marked "emergency exit" to leave the unwanted action without having to go through an extended process. | *When it's easy for people to back out of a process or undo an action, it fosters a sense of freedom and confidence. Exits allow users to remain in control of the system and avoid getting stuck and feeling frustrated.* |

[74] Nielsen & Molich, 'Heuristic evaluation of user interfaces'.
[75] Nielsen, 'Heuristic Evaluation', 25–62.

| | Title | Criteria | Impact |
|---|---|---|---|
| 4. | **Consistency and standards** | Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform and industry conventions. | *Jakob's Law[76] states that people spend most of their time using digital products other than yours. Their experiences with those other products set their expectations. Failing to maintain consistency may increase the user's cognitive load by forcing them to learn something new.* |
| 5. | **Error prevention** | Good error messages are important, but the best designs carefully prevent problems from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action. | *There are two types of errors: slips and mistakes. Slips are unconscious errors caused by inattention. Mistakes are conscious errors based on a mismatch between the user's mental model and the design.* |
| 6. | **Recognition rather than recall** | Minimize the user's memory load by making elements, actions, and options visible. The user should not have to remember information across interfaces. Information required to use the design (e.g., field labels or menu items) should be visible or easily retrievable when needed. | *Humans have limited short-term memories. Interfaces that promote recognition reduce the amount of cognitive effort required from users.* |
| 7. | **Flexibility and efficiency of use** | Shortcuts – hidden from novice users – may speed up the interaction for the expert user, such that the design can cater to both inexperienced and experienced users. Allow users to tailor frequent actions. | *Flexible processes can be carried out in different ways, so that people can pick whichever method works for them.* |
| 8. | **Aesthetic and minimalist design** | Interfaces should not contain information that is irrelevant or rarely needed. Every extra unit of information in an interface competes with the relevant units of information and diminishes their visibility. | *This heuristic doesn't mean you have to use a flat design. It's about making sure you're keeping the content and visual design focused on the essentials. Ensure that the visual elements of the interface support the user's primary goals.* |
| 9. | **Help users recognize, diagnose, and recover from errors** | Error messages should be expressed in plain language (no error codes), clearly indicate the problem, and constructively suggest a solution. | *These error messages should be presented with corresponding visual elements that will help users notice and recognize them.* |
| 10. | **Help and documentation** | It's best if the system doesn't need any additional explanation. However, it may be necessary to provide documentation to help users understand how to complete their tasks. | *Help and documentation content should be easy to search and focused on the user's task. Keep it concise, and list concrete steps that need to be carried out.* |

[76] See https://www.nngroup.com/videos/jakobs-law-internet-ux/.

# Annex II: ISEAL Credibility Principles

**'ISEAL Credibility Principles'**
These principles were originally developed in 2013 and updated in 2021 by the ISEAL Alliance.[77] Given the increasing volume of sustainability systems, the need for an international reference point to identify credible and effective systems was imperative. The principle number appears in column one, the title in column two, the criteria in column 3, and the impact in column four.

Table 17. 'ISEAL Credibility Principles' criteria and impact

| | **Title** | **Criteria** | **Impact** |
|---|---|---|---|
| 1. | **Impacts** | A credible sustainability system makes an impact where it matters. | *A credible sustainability system has a clear purpose to drive positive social, environmental, and economic impacts and to eliminate or remediate negative impacts. It defines and clearly communicates its scope, its specific sustainability objectives, and its strategies for achieving these objectives (its theory of change). The system focuses on the significant sustainability impacts in its scope. It seeks to address the root causes of sustainability issues and deliver wider or systemic impacts. It reflects current scientific evidence and international norms when relevant. It is adapted to local or sector-specific conditions where this helps improve impact.* |
| 2. | **Collaboration** | A credible sustainability system works with others to create change. | *A credible sustainability system identifies governments, private industry, and civil society organizations, including other sustainability systems that are working towards shared sustainability objectives. It actively seeks alignment and respectfully pursues collaboration with others. It establishes partnerships and shares learning to improve its efficiency and its direct or systemic impacts.* |

[77] ISEAL Alliance, 'ISEAL Credibility Principles'; see https://www.isealalliance.org/defining-credible-practice/iseal-credibility-principles.

| | Title | Criteria | Impact |
|---|---|---|---|
| 3. | **Value creation** | A credible sustainability system adds value. | *A credible sustainability system strives to create value that fairly rewards the effort and resources that it takes for users to participate in the system. It has a viable business model, and it operates efficiently, minimizing costs for users and reaching more users by reducing other barriers to access. It supports users to implement its tools, and it empowers users by demonstrating a clear business case for participating in its system.* |
| 4. | **Measurable progress** | A credible sustainability system can demonstrate the difference it is making. | *A credible sustainability system has tools that are relevant to achieving its sustainability objectives, and these tools allow progress towards objectives to be measured over time. It collects and analyzes the data it needs to measure, understand, and demonstrate the progress its users are making towards these objectives.* |
| 5. | **Stakeholder engagement** | A credible sustainability system listens and learns. | *A credible sustainability system is inclusive and non-discriminatory. It empowers stakeholders to participate in decisions and hold the system to account. It involves a balanced and diverse group of stakeholders in decisions that will affect them. It strives to understand the context and perspectives of stakeholders who have been under-engaged or under-represented, and it creates opportunities to ensure their participation in decision-making. It provides clear and transparent feedback on stakeholder input and concerns. It has fair, impartial, and accessible mechanisms for resolving complaints and conflicts.* |
| 6. | **Transparency** | A credible sustainability system earns trust by being open and honest. | *A credible sustainability system makes important information publicly available and easily accessible, while protecting confidential and private information. It enables stakeholders to understand and evaluate the system's processes, decision-making, results, and impacts. Stakeholders have the information they need to actively participate in decisions or raise concerns.* |
| 7. | **Impartiality** | A credible sustainability system is impartial. | *A credible sustainability system identifies and avoids or mitigates conflicts of interest throughout its governance and operations, particularly when it comes to assessing its users' performance. Transparency and stakeholder engagement help ensure the system's integrity can be trusted.* |

| | Title | Criteria | Impact |
|---|---|---|---|
| 8. | **Reliability** | A credible sustainability system provides trustworthy assessments of users' performance. | *A credible sustainability system designs its tools so that these can be consistently implemented and assessed. It ensures assessments of users' sustainability performance are competent and accurate, and that these assessments support any claims it allows users to make.* |
| 9. | **Truthfulness** | A credible sustainability system's claims and communications can be trusted. | *A credible sustainability system substantiates its claims. Any claims the system or its users make are clear, relevant, and can be checked. They enable customers and other stakeholders to make informed choices. The scope and design of the system is accurately reflected in any claims, ensuring these are not misleading. Claims about sustainability impacts are backed up with data and evidence that is publicly available.* |
| 10. | **Continuous improvement** | A credible sustainability system keeps improving. | *A credible sustainability system regularly reviews its objectives, its strategies, and the performance of its tools and system. It evaluates the impacts and outcomes of its activities. It applies the lessons learned to improve. It responds to new evidence, stakeholder input, and external changes, adapting its strategies to improve its impacts and remain fit for purpose.* |

# Author

**Hadley Newman** is a strategic communications advisor working across a broad range of policy areas for public and multilateral organizations. Formerly strategic communications lead on the G20 and a senior director with a global public affairs consultancy, he specializes in strategic communications planning, and the empirical analysis of target audiences. Mr Newman is a published author on communication governance and foreign information manipulation and interference. His doctoral research at Heriot-Watt University explores the influence of targeted communication within information operations.

Hybrid CoE

The European Centre of Excellence
for Countering Hybrid Threats