Global Operations Primer - Health Misinformation

NB: This is intended to provide a high-level overview of the Global Operations organization, but tailored to context for cross-functional partners involved in COVID Defense efforts on Misinformation. Our support for COVID Defense is continuously changing, and so we will work to keep this document up to date. Please reach out to @Alexis Link or @Cristina Cepero-Novoa with any questions or feedback on how to improve this or if you need to make edits.

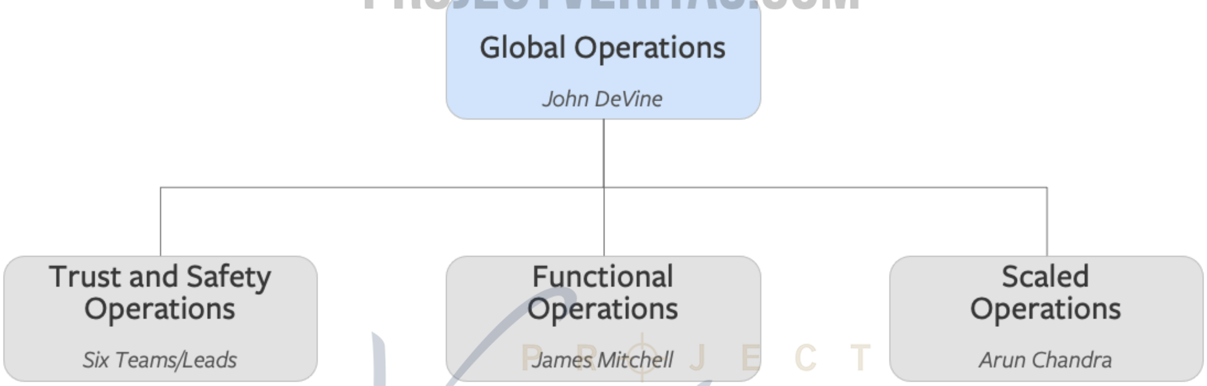
Last Updated: *April 12th*

Organizational Structure and Purpose

The mission of Global Operations (GO) is to:

> *Build and run world-class processes at a global scale that minimize harm to people and society, and maximize success and well-being of our ecosystem of people, community, business, and partners.*

To achieve this mission, the organization is split into three key pillars under GO's leader, John DeVine.



Very broadly speaking, Trust and Safety is responsible for *demand* (or what work our reviewers do), and Scaled Ops is responsible for *supply* (the people who do that work). Functional Operations covers Risk and Response and houses the GO Central Analytics team and GO Engineering teams. These three functions are very different in their scope, responsibilities, and requirements.

As a general rule, all **scaled review**–or enforcement done by our outsourced review teams across many languages and at high volume–is operationalized by someone within Trust and Safety, partnering with Scaled Ops to successfully launch. These flows are almost always in SRT. On the other hand, **escalations-only** or specialized flows are owned by Risk and Response teams (Functional Ops) or Market FTEs (Scaled Ops), and may or may not be in SRT.

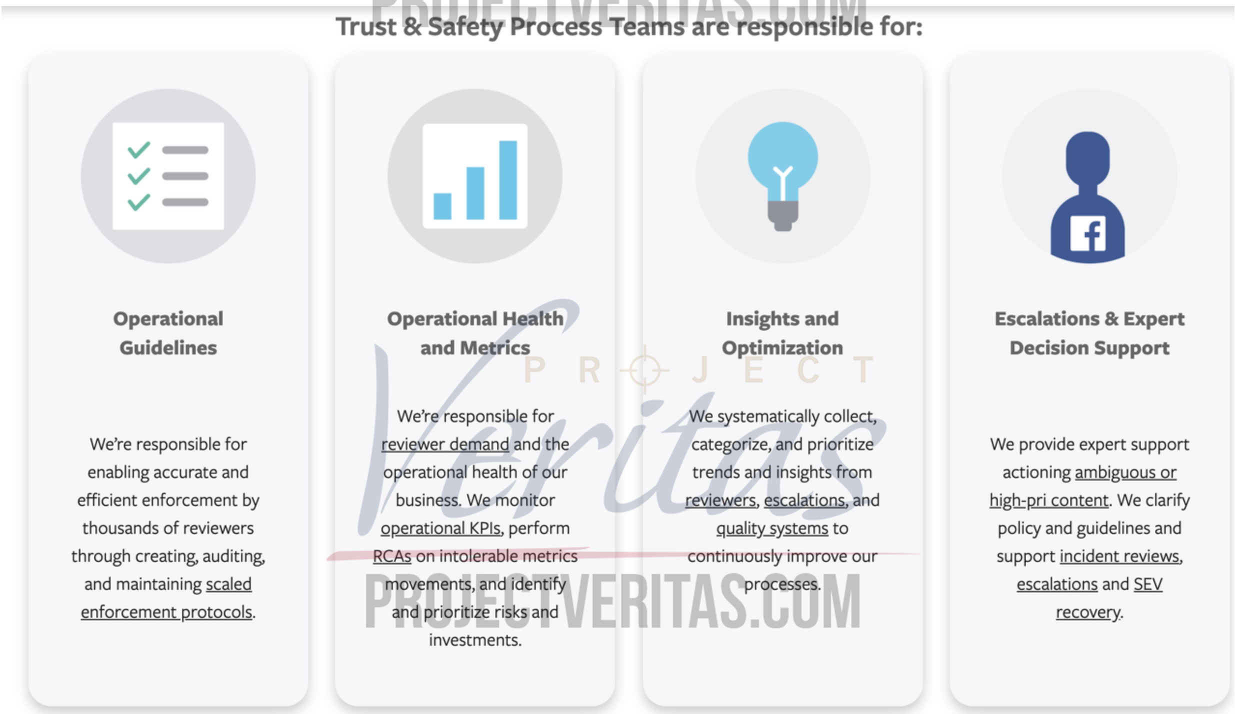
		What?	Reviewers	Tools
1	Scaled Review	High volume review across all supported	Outsourced Reps	SRT

1	Scaled Review	languages	Outsourced Reps	SRT
2	Escalations Only	Small-volume review on particular high-priority areas and escalations-only policies	FTEs, sometimes Contractors	SRT, Centra, Tasks, other

Trust and Safety is made up of a number of separate teams depending on the function of the business:

1. Community Operations (CO) covers enforcement, measurement, and labeling for the **Implementation Standards** (internal-facing guidelines for how to enforce the Community Standards) for **organic content;** the Misinformation Process team sits within CO
2. Product Data Operations (PDO) lives within Product Support Ops and covers **Product Policy,** or the guidelines dictating how people can actually use our products (e.g. Groups or Live policies, platform policy, etc.). Product Policy is policy that applies to a particular experience or product experience (for example Fundraisers or Dating) to address risks that are unique to those surfaces which can't be addressed through Community Standards or product controls.
3. Business Integrity (BI) covers enforcement for **ads policies**
4. Commerce Ops is responsible for **connections between businesses and users**
5. Risk and Payments (R&P) protects the community from **financial abuse** and provides **payments support**
6. Legal and Premier Partner Operations supports **rights holders, media partners,** and their content

Though the focus areas differ for each of the Trust and Safety teams, their responsibilities are generally similar.



Trust & Safety Process Teams are responsible for:

Operational Guidelines

We're responsible for enabling accurate and efficient enforcement by thousands of reviewers through creating, auditing, and maintaining scaled enforcement protocols.

Operational Health and Metrics

We're responsible for reviewer demand and the operational health of our business. We monitor operational KPIs, perform RCAs on intolerable metrics movements, and identify and prioritize risks and investments.

Insights and Optimization

We systematically collect, categorize, and prioritize trends and insights from reviewers, escalations, and quality systems to continuously improve our processes.

Escalations & Expert Decision Support

We provide expert support actioning ambiguous or high-pri content. We clarify policy and guidelines and support incident reviews, escalations and SEV recovery.

Health Misinformation Support

Health Misinformation *scaled* review is supported by both CO and PDO:

1. PDO is *labeling* to train Barriers to Vaccination (B2V) classifiers;
2. CO outsourced reviewers are *enforcing* on Misinformation and Harm (M&H), Widely Debunked Hoaxes (WDH), and Repeatedly Fact-Checked Hoaxes (RFH), conducting prevalence labeling, supporting appeals, and conducting quality reviews

CO scaled review is done by ring fenced reviewers–meaning these reviewers are **only** working in the health space. This is due to the extremely complex nature of misinformation policies and current challenges in accuracy. However, this limits flexibility of the workforce; we can't simply add more work to the reviewers as all new work must tradeoff on other health work. We are actively exploring avenues to break the ring fencing model; this comes with other challenges, and is unlikely to happen before the end of Q3 at the absolute earliest.*

This scaled review is currently in 10 languages covering 90% of MAP: English, Spanish, Portuguese, Hindi, Bengali, French, Arabic, Thai, Indonesian, Tagalog, Italian, Burmese, Dutch, German, Malay, Polish, Turkish, Ukrainian, Vietnamese

This scaled review only covers simple objects–meaning posts, comments, photos, videos, etc. We are planning to expand coverage to complex objects (pages, groups, events, IG profiles, FB profiles) in late Q3.* Complex Object review is significantly more resource-intensive than simple objects, and therefore reviewers get through fewer per hour. Further, all enqueued content is proactively detected–we do not have user reporting today for misinformation.

** Breaking the ring fenced model and expanding to complex objects will take until at least Q3. This is due to a number of factors, including but not limited to challenges in hiring due to COVID affecting cross-border work, complexity of misinformation policies, current accuracy, reviewers using an "Info First" review tree versus standard three-step labeling tree, and developing new tooling.*

Non-Scaled Support

Markets
We also have support from our Market FTE teams. An important note is that these FTEs are *not* an enforcement workforce. Unlike our CO outsourced reviewers, who spend ~24.5 hours per week on content moderation, Market FTEs only spend a part of their time doing content review.

Each week, we have Market FTEs enforce for a few hours in the following languages: Amharic, Georgian, German, Malay, Persian, Turkish, Russian, Albanian, Croatian, Czech, Hungarian, Kurdish, Maghreb, Slovak, Afrikaans, Finnish, Nepali, Norwegian, and Zulu. We are also requesting regular support for the languages supported at outsourcing to help on complex objects (or entities). Further, Market FTEs across all markets staff X-Check queues and escalations that come in.

Unlike CO outsourced reviewers, FTEs can enforce on all policies–including escalations-only policies. This covers M&H, WDH, RFH, B2V tiers 1 and 2, and Dedicated Vaccine Discouraging Entities (DVDE). They primarily enforce on simple objects which are sourced from Health Integrity managed classifiers and GO Markets managed CIRD pipelines, but, by covering the DVDE policy, will also begin to support complex objects.

Within the Markets team sits the Civic Incubator team. This is made up of FTEs from the North America market. This team covers COIL (Complex Objects Integrity Lab), formerly known as HEROCO, which is an SRT flow allowing for in-depth review of English Groups (and soon Pages and IG Profiles) against all policies. Today, this review process takes about 10 minutes per Group reviewed. In addition, this group will soon start review and enforcement of escalated COVID Top 100 content and entities.

Finally, we have Market FTEs and contractors who work on HERO. HERO reviewers cover the predicted top 1000 posts in the US by VPVs daily. These reviewers apply neutral inform treatments, or NITs, within the HERO flow. They also temporarily supported COVID Top 100 post review for one week.

Appeals from Markets work outside of COIL currently route to normal CO scaled review. As a result, the reviewers for appeals are not trained in all policies. We are discussing a proposal to prevent appeals for complex objects (entities) for reviews done by Market FTEs.

Risk and Response

There are two risk teams and six escalation teams within Risk and Response. The three key ones for Misinformation *enforcement* are CO-PREsc, Early Response Escalations, and Regulation Escalations (the teams that enforce on organic content). These three teams are responsible for enforcement as *escalated* by internal staff, governments and NGOs, M-Team, and the Media. They enforce on every surface and product within the Facebook ecosystem, with the exception of WhatsApp. These teams are responsible for all Implementation Standards, including the 60+ escalations-only policies and MisInformation policies. Beyond enforcement, these teams are lead crisis managers (e.g. Capitol Riots). In addition to staffing all L2+ L3 IPOCs, senior members of these teams serve as IPOC Crew Leads, responsible for IPOC effectiveness and communications. As a result, these teams have very little bandwidth for additional support.

These teams have global presence, which facilitates 24/7 coverage. That said, they are relatively small teams (5-15 per region per team) with a dynamic workflow, meant to pivot into and out of crisis, and only support English content.

Response teams work in Centra instead of SRT, use Cases as their case management tool (as opposed to tasks), and have access to most tooling available (including BAT, MMS, CET, BH, and XCheck).

Escalation teams are organized by escalated content source, i.e. government/NGO escalations, media, Oops, or book of business partner. The teams are not trained in B2V tiers 3 and 4, as these will be automated only. In addition, they will enforce on COVID-topic Top 100 FB and IG pages, groups, IG profiles, and comments flagged by First Responders until this flow is passed off to Civic Incubator. We do not offer appeals for decisions made by PREsc.

First Responders are a contractor team under PREsc. They will be filtering through some Top 100 content and entities to be escalated to PREsc or Civic Incubator.

Outside of enforcement, the Community Risk Assessment team is responsible for the Dynamic Risk Assessment (DRA) shared weekly, proactive risk investigations (PRIs), and the weekly incident review.

Health Misinformation Enforcement Summary

	Workforce	Ops POC	Employee Type	Policies	Review Type	Languages	Job Types	Scale
1	PDO	@Fiona Yee	Outsourcing	B2V	Labeling	Many	Content	High
2	CO Ring Fenced	@Reshama Deshmukh	Outsourcing	M&H, WDH, RFH	Enforcement, Measurement	19	Content	High
3	Markets - General	@Orla Power	FTE	M&H, WDH, RFH, DVDE, B2V 1/2, CH	Enforcement	~25	Content, Entities	Medium
4	HERO	@Dylan Ackerman	FTE, Contractor	B2V (Inform treatments only), M&H, RFH, WDH	Enforcement, Inform Treatments	~25	Posts	Medium
5	COIL (owned by Civic Incubator)	@John Shea	FTE	M&H, WDH, RFH, DVDE, B2V 1/2, CH	Enforcement	English	Groups	Medium
6	Civic Incubator	@Tori Manlove	FTE	M&H, WDH, RFH, DVDE, B2V 1/2, CH	Enforcement	English	Content, Entities	Medium
7	PREsc	@Caroline Nichols	FTE	M&H, WDH, RFH, DVDE, B2V 1/2, CH	Enforcement	English	Content, Entities	Low
8	Early Response	@Zach Gerasin	FTE	M&H, WDH, RFH, DVDE, B2V 1/2, CH	Enforcement	English	Content, Entities	Low
9	First Responders	@Caroline	Contractor	M&H, WDH, RFH,	Enforcement,	English	Content, Entities	Low

9	First Responders	Nichols	Contractor	DVDE, B2V 1/2, CH	Labeling	English	Content, Entities	Low

Policies Glossary

1. **M&H:** Misinformation and Harm
2. **WDH:** Widely-Debunked Hoaxes
3. **RFH:** Repeatedly Fact-Checked Hoaxes
4. **CH:** Coordinating Harm
5. **DVDE:** Dedicated Vaccine Discouraging Entities
6. **B2V:** Barriers to Vaccination
7. **VH:** Vaccine Hesitancy