



Building a safer network, together

jim@foundation.matrix.org
@macksque:matrix.org

The Foundation Safety Team is...

an **operational** team

an **engineering** team

a **standards** team

a **policy** team

a **research** team

a **small** team, funded by you

7,300,000+

The number of events processed by the Foundation policy server in the last 12 months

20,000+

The number of reports from matrix.org users the team reviewed in 2025

99.9%

The accuracy rate of the team in QA checks across the year

Reasons to be cheerful

Celebrate every win

Better **governance**



Trust & Safety Committee

Trust & Safety Research & Documentation Working Group

Safety Guild

Room Directory Working Group

Easier access to safety tools



Asgard.chat

Meowlnir policy server

Foundation Policy Server

Draupnir: zero touch deployments

Open registration policy list

hma-matrix

Continuuity: safer defaults on open registration

ROOST

Matrix v1.18

- account locking
- account suspension
- invite blocking

More **cooperation**

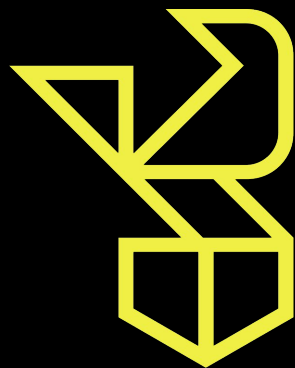


Muninn Hall a community for public federation server staff

Safety Guild a working group for people building safety tooling for Matrix

Community Moderation Effort (CME) a shared policy list combating spam

ROOST

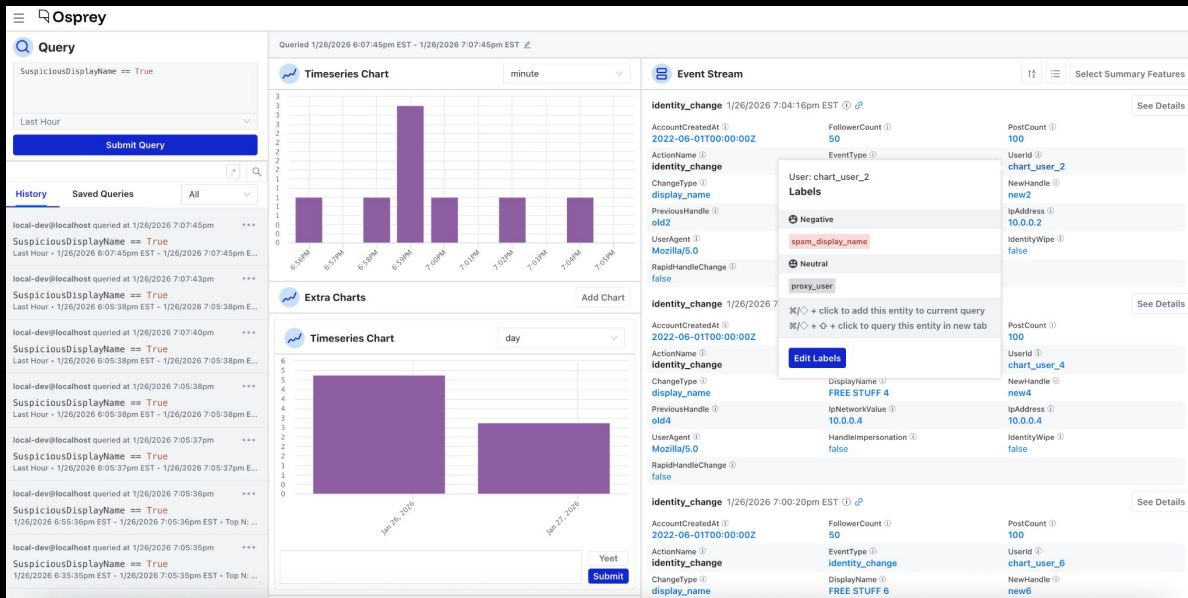


roost

 Coop +  Osprey



- In-depth explorations of potentially violating behaviors
- High queries-per-second (QPS) for streaming and batched data
- Open-ended investigation
- UI for analysts to identify abuse patterns and signals





- Flexible review tool for labeling multiple formats (ie. content and accounts)
- Automated routing of tasks into queues
- Queue orchestration, audit trails
- Appeal workflows
- Reviewer wellness features
- Integrated with HMA for hash matching and Content Safety API for novel CSAM detection

Review Console

Review queues allow you to manually review content and users one at a time

Create Queue

Filter

ID	Name	Description	Pending Jobs	
8b994b20-ecf4-11f0-8660-35b4d4233032	Media	images and videos only	13	<button>Start Reviewing</button> <button>Delete All Jobs</button>
848799e0-ecf4-11f0-8660-35b4d4233032	Default Queue	default fallback queue	20	<button>Start Reviewing</button> <button>Delete All Jobs</button>
7a055bd0-f08a-11f0-83c4-8ffe1027355f	Fraud	Flagged by signals indicating fraud	0	<button>Start Reviewing</button> <button>Delete All Jobs</button>
fab56b40-f31d-11f0-8e8b-f366a41f4608	Scams	User reports or signal-based likelihood of scams and social engineering	0	<button>Start Reviewing</button> <button>Delete All Jobs</button>
33288f50-f235-11f0-9ca1-49f74809974b	Spam	Suspected spam behavior or posts	0	<button>Start Reviewing</button> <button>Delete All Jobs</button>

Looking **forward**

Coop + Osprey



Now

In production for matrix.org, alongside legacy systems

In 2027

Switch off legacy systems

Open source **Raven**

Share our configs for matrix.org

polycserv

Community Guidelines

reporting v2

Where we **are**

[**matrix**]

It's... functional.

Where we are **going**

- + Reporting over federation
- + Reporting to community safety teams
- + Appeals
- + Improved clarity on reasons for enforcement
- + Shared language of harms
- + Resources for users
- + Transparency reporting
- + Reporting the content of encrypted events

MSCs



MSC4387

M_SAFETY error code

MSC4414

Design decision - Errors

MSC4456

Harms taxonomy

MSC4457

Generic reporting API

MSC4468

Reporting to communities (via to-device)

MSC4469

Reporting to remote servers (EDU approach)

MSC4470

Routing reports to non-local destinations

How **you** can help

time + attention

money

celebrating **wins**



Thank you for listening

Please find me for questions
today or tomorrow

Useful Links



Tools

[Draupnir](#), a community management platform and moderation tool

[Meowlnir](#), a moderation bot & policy server

[Mjolnir](#), a moderation bot

[Asgard.chat](#), hosted versions of Draupnir & Meowlnir

[Coop](#), a review and moderation tool from ROOST

[Osprey](#), a safety rules engine and investigation tool from ROOST

[policyserv](#), a community-centric policy server for Matrix

[hma-matrix](#), extensions to [HMA](#) to make it work for Matrix

[Ketesa](#), a tool for server administration

[Supervision](#), a moderation tool in Element Server Suite Pro (**ESS Pro**)

Communities

[Muninn Hall](#), a community for public federation server staff

[Safety Guild](#), a working group for people working on safety tooling

[Community Moderation Effort \(CME\)](#), a community spam blocklist

[ROOST](#), a non-profit building open source safety tools

[T&S Research + Documentation WG](#), working on guides and documentation

Resources

[Open registration server list](#), a policy list of unsafe servers