# Democratizing and decentralizing the DSA

Onno Hansen-Staszyński
E: onno@drog.group

Global Age Assurance Standards Summit 2026
Manchester, 14.4.2026

Saufex

This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement Nº 101132494.

# Project SAUFEX - objectives

- Standardise knowledge, building on existing good practices to further develop a common and standardised framework for FIMI analysis;

- **De-centralise and democratise processes around FIMI analysis and response;**

- **Incorporate community-driven quality assurance processes** and expand theoretical understandings of FIMI within the defender community.

Saufex

This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement Nº 101132494.

# DSA de-centralisation and democratisation – mass-flagging app GetResilience

- Inclusiveness

- Supports **DSA enforcement**
- Enables **systemic risk analysis**
- Ensures **GDPR compliance**

**Dual output:**

- Legal action
- Risk intelligence

Saufex

This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement Nº 101132494.

# DSA de-centralisation and democratisation isn't obvious - dilemmas

- The app should have **low entrance friction** for end-users who wish to flag but at the same time result in **useful flags**, that is (1a) flagging illegal content or (1b) content or procedures that are considered by the flagger to be harmful content;
- The app should have **low entrance friction** for end-users who wish to flag but the app could face **penalties for abusive reporting**;
- The app should have **low entrance friction** for end-users but should not be a target for **adversarial gaming**, e.g. "mass-flagging" by bots or coordinated actors to censor legitimate speech;
- The DSA process requires a **good faith statement** but the app needs to allow for anonymous **flagging**;
- The DSA process requires **a good faith statement** which makes **AI automation** of the flagging flow that would enable mass access to flagging a challenge;
- The DSA requires **flagger accountability and trustworthiness** but that triggers **GDPR** risks;
- Non-anonymous flagging potentially triggers the **GDPR** but the app should be as **automized** as possible;
- Regular end-users are not trained in legal details but insufficient knowledge of the details potentially **disqualifies** the flag: e.g. reporting non-illegal FIMI as "illegal content" leads to "liability leakage" and regulatory rejection;
- The **DSC** is to be informed at an earlier stage and more comprehensively but the DSA flags are **platform-oriented**;
- The DSA is **platform-oriented** but FIMI and other threats are often **cross-platform campaigns**.

Saufex

This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement Nº 101132494.

# GetResilience, draft core architecture

**Two completely separate tracks:**

- **Track A:** Legal notices (illegal content only)
- **Track B:** Risk intelligence (non-legal)

**Key rule:**

- No mixing between tracks

Saufex

This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement Nº 101132494.

# User types

- Anonymous or identified users – random individuals entering simple flagging input as signal (pre-track input)

- Formal members organized in so-called Resilience Councils; they create legal outputs (Track A) or validate threats (Art. 18)

- Owner of the system - a formal organization. The formal organization is ultimately legally responsible for all outputs.

Saufex

This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement Nº 101132494.

# Resilience Councils

- Inspired by Polish RCs
- Formal members organized in groups of internally certified experts
  - EMoD online learning modules
- The groups are organized by violation domain and jurisdiction

Saufex

This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement Nº 101132494.

# General users

**Initiate the flow: step 1**

- User submission of
  - URL
  - Platform
  - Neutral classification
  - (For non-anonymous users) Context
  - (For non-anonymous users) Structured categorisation

System treatment:

- Stored as non-attributable signal
- No legal qualification
- No evidential status

Saufex

This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement Nº 101132494.

# Step 2 - system

System performs non-decisional and non-legal routing:

- Basic sorting:
  - Priority level
  - Violation domain
  - Jurisdiction cluster
  - Risk indicators
- Routing to:
  - Standard review queue (there is no potential threat to life/safety)
  - or priority/threat queue (there is a potential threat to life/safety)

Saufex

This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement Nº 101132494.

# Step 3a - rapid threat validation (Art. 18)

- Actor: RC member
- Activity: Performs rapid human assessment
- Applies: “Reasonable suspicion” standard
- If threshold met:
- Direct report to: National law enforcement or Europol
- If threshold NOT met:
  - Signal downgraded
  - Returned to standard flow

Saufex

This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement Nº 101132494.

# Step 3b - verification

- Actor: RC member
- Mandatory actions:
  - Independently access content
  - Verify existence and context
  - Reconstruct evidence outside the system
  - Apply legal reasoning
- Critical rule: User input is not relied on as evidence
- Decision: Does the RC member form a reasonable legal assessment of illegality?
  - If no: to track B
  - If yes: track A

Saufex

This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement Nº 101132494.

# Step 4 – track A

- RC member drafts legal notice (not yet valid) - legal qualification, supporting reasoning, evidence basis – and send it for verification

- Two additional RC members independently review and confirm reasoning
  - If consensus reached: Notice is: co-signed by 3 RC members - it becomes a formal notice only after consensus - submitted to platform (Art. 16 DSA)
  - If consensus fails: case dropped or reworked

Saufex

This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement Nº 101132494.

# Step 5 – track B

- For:
  - Harmful but legal content
  - Failed legal threshold cases
- Processing:
  - Aggregation
  - Pattern detection
  - Cross-platform analysis
- Output:
  - Anonymised intelligence
  - Strict constraints:
  - No attribution
  - No enforcement
  - No legal claims

Saufex

This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement Nº 101132494.

# End States

**Law enforcement report (Art. 18)**

- Rapid escalation
- Based on “reasonable suspicion”
- Independent of Track A

**Legal notice (track A)**

- Fully human-authored
- Legally attributable
- Sent to platform

**Intelligence output (track B)**

- Aggregated insights
- No legal effect

Saufex

This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement Nº 101132494.

# Age verification

- Age verification is not relevant for VLOSEs.
- It may be relevant for VLOPs.
  - Introducing age verification for flagging on VLOPs would be formally aligned with bans on minors' access in certain jurisdictions.
- Under Art. 24(2) CFR, the child's best interests must be a primary consideration.
  - If minors fully complied with the bans, age verification for flagging would be redundant.
  - In practice, a non-trivial number of minors will circumvent such bans and access VLOPs – e.g. Australia 3 months in
  - For those minors, the relevant question is whether their best interests are served by denying or providing access to harm-mitigating tools such as flagging mechanisms.
  - The inclusion of age verification therefore depends on whether exclusion from such tools is in the child's best interest.

Saufex

This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement Nº 101132494.

# No age verification

- Recital 89 DSA provides guidance, stating that VLOPs should ensure their services are organised so that minors can easily access mechanisms such as notice-and-action, “where applicable.”
  - While “where applicable” may appear to allow exclusion, this cannot be relied upon when the presence of minors and their exposure to harm are foreseeable. The DSA is risk-based rather than purely status-based, and foreseeable risks must be mitigated irrespective of formal access restrictions.
- Therefore, excluding minors from a harm-mitigating mechanism such as mass flagging would be difficult to reconcile with both Art. 24(2) CFR and the DSA’s risk-mitigation logic.
- For that reason, the mass-flagging tool will not include age verification, as enabling access to such mechanisms better serves the child’s best interests through risk mitigation.

Saufex

This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement Nº 101132494.

# Further reading