



ReCreating Europe



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870626

Platforms' content moderation & accountability

Evaluating the present and looking into the future

Hosted by Christian Katzenbach & Sebastian Schwemer




Online Workshop (Copenhagen/Bremen)

8 December 2022





Some housekeeping

- Please keep your microphones off during the presentations 
- Slides will be uploaded on Zenodo after the workshop 
- Roundtable discussion (rather than presentations) 



ReCreating Europe

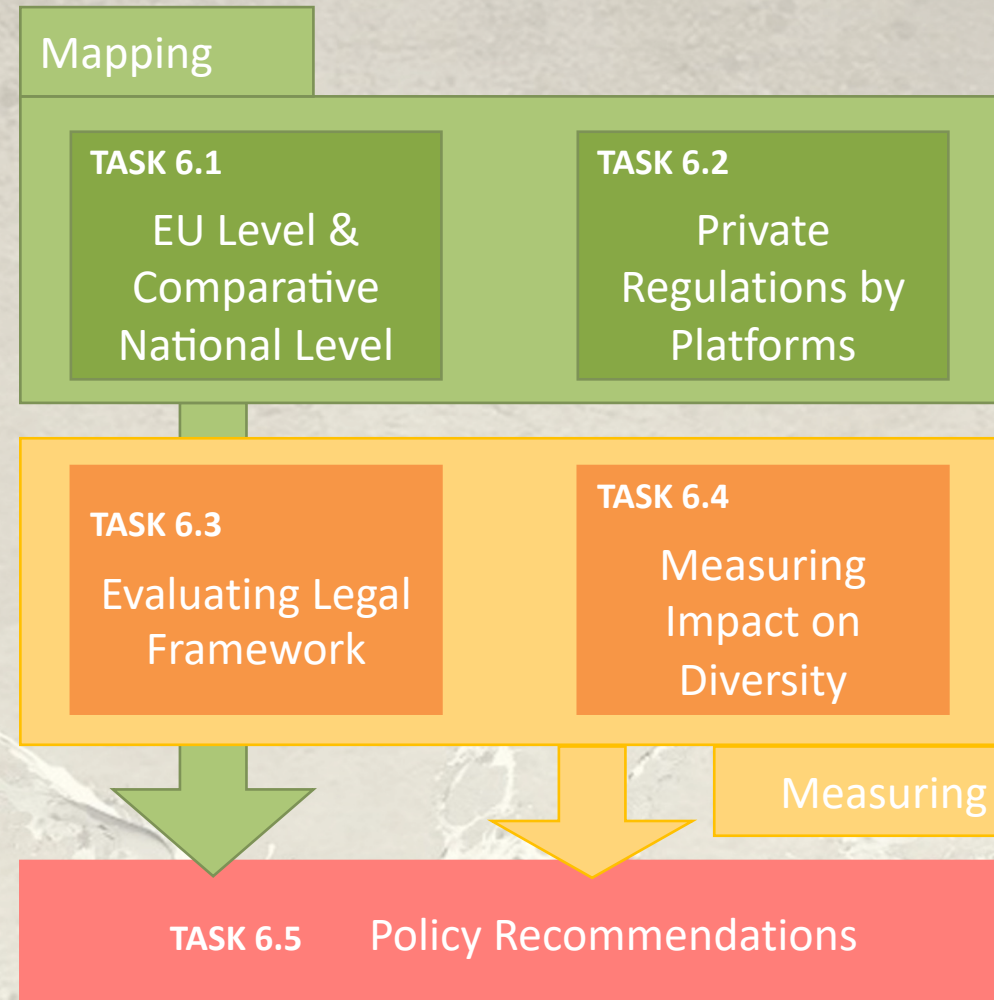
ReCreating Europe

COPYRIGHT CONTENT
MODERATION IN THE
INTERDISCIPLINARY M
ANALYSIS

Impact of content
moderation practices
and technologies on
access and diversity

D.6.3. Final Evaluation and Measuring Report
Authors
Sebastian, Christian, Thomas, Daria, João

João Pedro Quintais, Péter Mezei, István Magalhães, Christian Katzenbach, Sebastian and Thomas Riis
August 2022





Two starting points:

- **Part 1: Overlaps and missing pieces:** Post-DSM Directive and DSA, where do we go from here, what is missing?
- **Part 2: Measuring and transparency:** hurray for mandatory data access regimes for researchers – but (how) can these be operationalised for the study of content moderation?



Overlaps and missing pieces: Post-DSM Directive and DSA, where do we go from here, what is missing?



Normative evaluation of frameworks

A theory of „rough justice“ for internet intermediaries from the perspective of EU copyright law

Examine the role of bias + (training) data in © content moderation (and recommendation)

Quality of Automated Content Moderation: Regulatory Strategies for Mitigating Error

access to data...? (and how to operationalise it?)



Quality of automated content moderation

- Benchmark for decision quality?
 - Substantive legal rules/private rules *inter partes*/users' ~~normative~~ ~~perception~~
 - Assumption: "quality" of copyright CoMo is correlated to access to culture (considered embedded in the existing copyright framework)
- "quality": in simple terms correct and false results. But: **What error rate is acceptable under the legislative framework?**
 - DSA: „Accuracy“ in DSA reporting; „sufficiently reliable to limit to the maximum extent possible the rate of errors“ (error rate not zero)
 - CDSMD: 17(7) CDSM „shall not result in the prevention“; 17(9) para. 3 CDSM „shall in no way affect legitimate uses“; EC's Guidance on Article 17 "to restore legitimate content ex post" would "not be enough for the transposition and application of Article 17(7)" → limit to manifestly infringing uploads; AG Øe: „negligible number of cases of 'false positives'“ but error rates "should be as low as possible"...
- Ex post mitigation mechanisms; but ex ante...?

Copyright infringing		
yes	no	
True positive (TP)	False positive (type-I error)	yes Takedown
False negative (type-II error)	True negative (TN)	no



A model of rough justice

- Procedural rules
 - need for more transparency into how content moderation works
 - appeal process in CoMo is not comparable to the traditional perception of fair trial and significant limitations in procedure must be accepted (e.g., evidence admissible, extent of evidence, number of pleadings)
- Substantive rules
 - Should create a counter-weight to internet platforms' tendency to over-enforce.
 - Should reduce moderation of incompatible but legal content.
 - Substantial rules based on human rights as means to align the platforms' ToS to societal objectives and value? Direct applicability of international human rights to platforms necessary?
- Competences of humans involved
 - Random test of accuracy by human intervention?
 - Adequate training and working condition? (Time spent on decision)

Measuring the impact of moderation practices and technologies on access and diversity

Transparency: hurray for mandatory data access regimes for researchers – but (how) can these be operationalised for the study of content moderation?



ReCreating Europe



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870626

Thank you for your participation!

Follow us:

[Facebook](#) | [LinkedIn](#) | [Twitter](#) (@reCreatingEU)

Contact us:

www.recreating.eu

