

VENDREDI 3 OCTOBRE 2025

HACK THE DSA

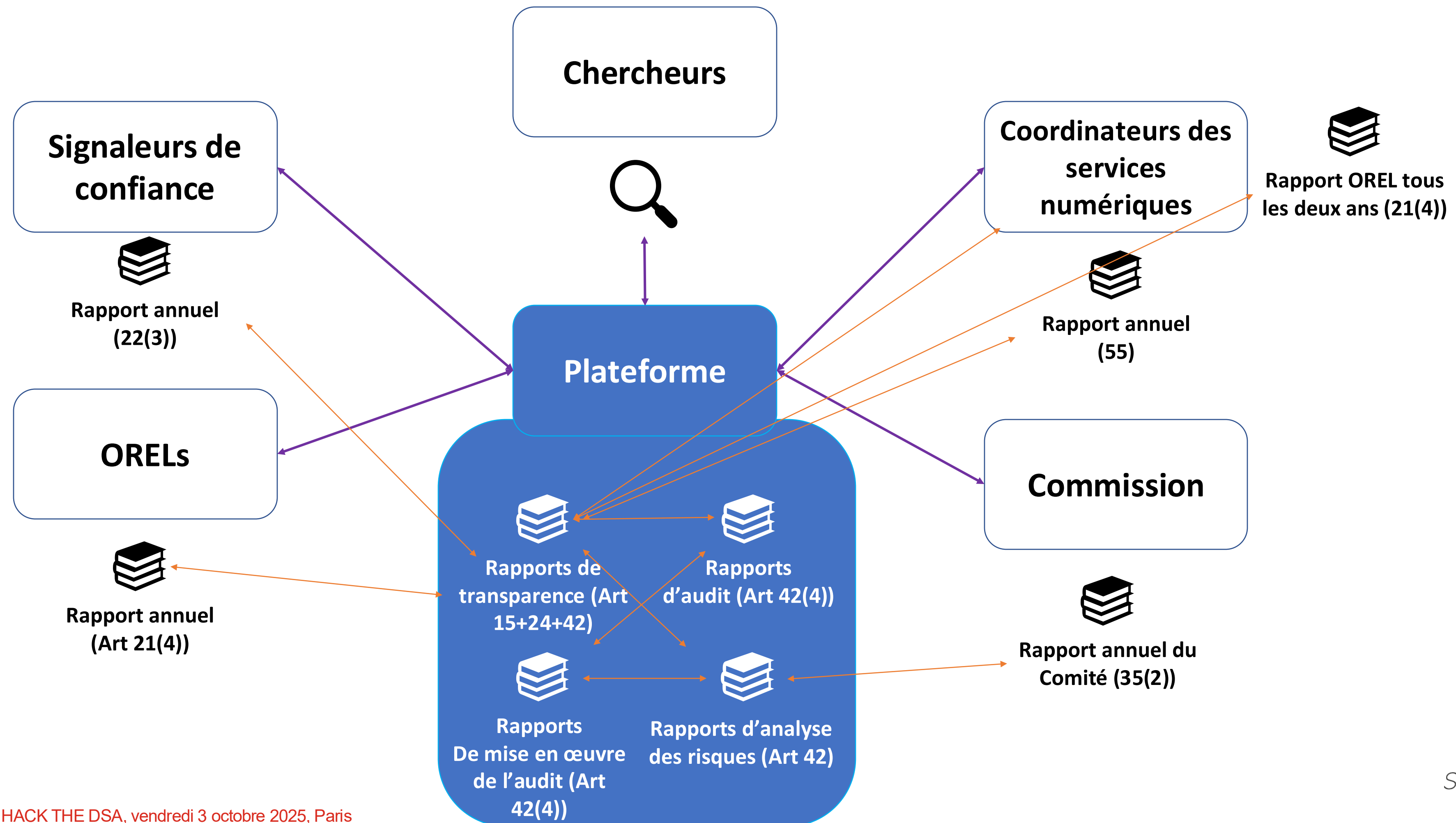
Inter-transparence



chaire 
sur la modération
des *contenus*

Constats et observations

Le DSA prévoit un large éventail de mesures de transparence isolées



Le défi de l'inter-transparence

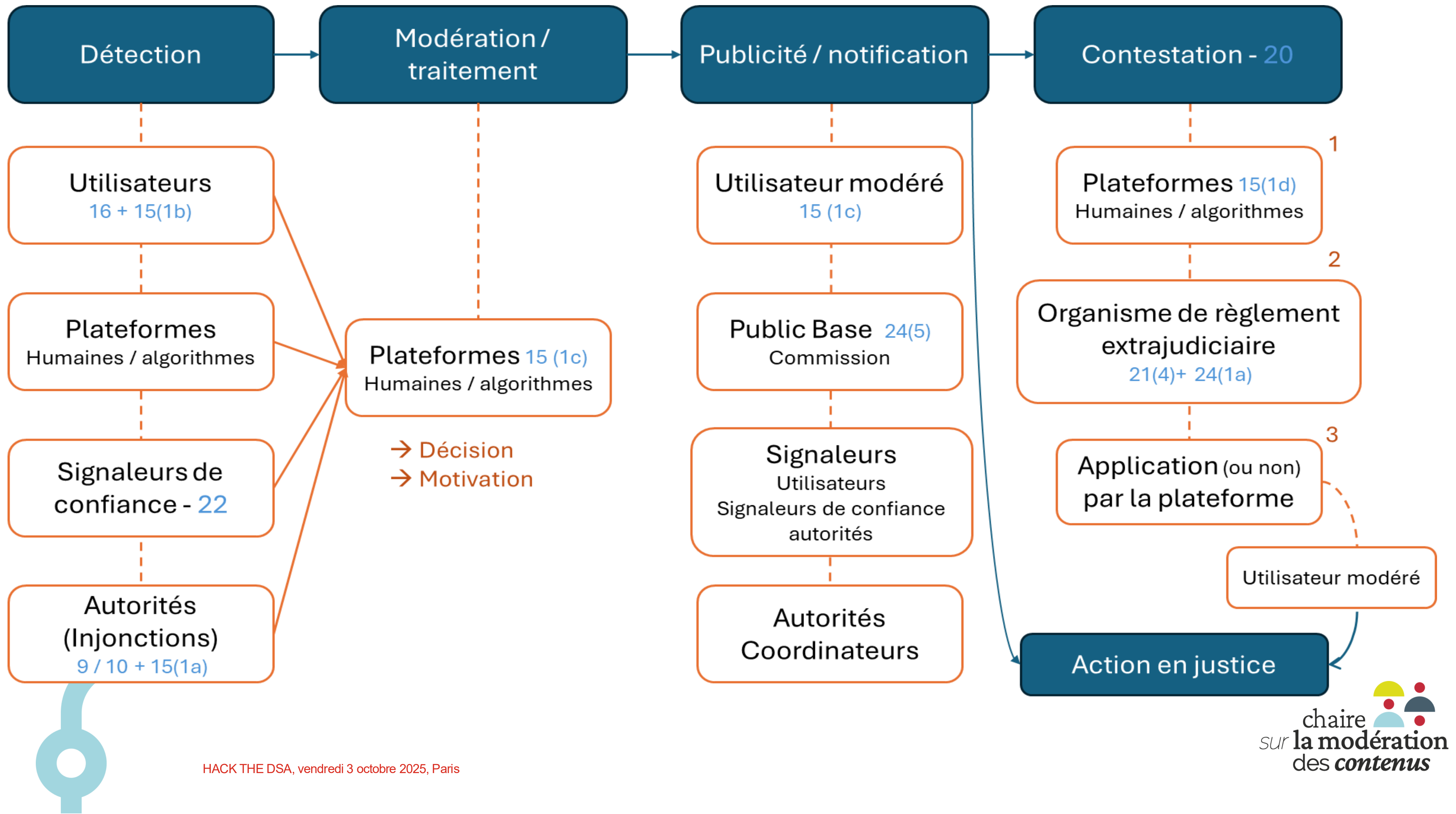
Impossibilité actuelle de comparaison absolue des données

Impossibilité de comparaison : catégorisation différente en fonction des types de rapports et des acteurs.

Eparpillement des données : l'absence de centralisation des données.

Comment tirer le meilleur profit de la transparence offerte par le DSA ? Et comment trouver une clef de lecture commune à toutes les données éparpillées ?







Détection

Sources interrogées

Art. 40

Art. 15

Signalements
utilisateurs

Signaleurs de
confiance

Autorités
(Injonctions)

Détection
proactive

Humaines/ algorithmes

Rapport de
transparence
de la
plateforme

Art. 22

Rapport des TF

Art. 55

Rapport DSC

Art. 34 et 35

Risques
systémiques
évaluation du risque
associé

Art. 37

Rapport
d'audit
comparaison
qualitative

Données des chercheurs*

*Échantillon aléatoire

Exemple

1 📁 Liste des dispositifs d'ouverture

	📄	Fondement pr...	Nom
1	📄 🔗	Art.39	Registres publicitaires
2	📄 🔗	Art.40 §12	Données des plateformes publiquement accessibles
3	📄 🔗	Art.40 §4	Données des plateformes non-publiquement accessibles
4	📄 🔗	Art.24 §5	Base de transparence des décisions de modération
5	📄 🔗	Art.42 §1	Rapports de transparence des très grandes plateformes en ligne
6	📄 🔗	Art.24	Rapports de transparence des plateformes en ligne
7	📄 🔗	Art.15	Rapports de transparence des services intermédiaires
8	📄 🔗	Art.42 §4	Rapports d'évaluation des risques
9	📄 🔗	Art.37	Rapports d'audit des très grandes plateformes
10	📄 🔗	Art.22 §3	Rapports d'activité des signaleurs de confiance
11	📄 🔗	Art.21 §4	Rapports d'activité des organes extrajudiciaires
12	📄 🔗	Art.55	Rapports d'activité des coordinateurs
13	📄 🔗	Art.35 §2	Rapports du comité européen

Grist
collaboratif



Limites identifiées

La catégorisation des contenus sera toujours relative

Surtout pour des contenus relevant du discours haineux, de la désinformation, etc.

Les chiffres ne seront pas parfaitement superposables

L'idée n'est pas de comparer au détail près (il y aura forcément des incohérences, notamment liées aux problèmes de qualification) mais **d'observer des tendances**.



Cas d'usage

- **Contrôler la cohérence globale des données.**
- **Permettre une comparaison avec des données issues de la recherche (Art. 40) :** l'objectif est d'enrichir le tableau avec des données issues de la recherche (des échantillons « témoin ») pour identifier certaines tendances ou certains biais (ex : surreprésentation d'un type de contenu dans les signalements par rapport à sa présence sur la plateforme ; mesure de l'efficacité de la modération automatisée, etc.)
- **Déduire des patterns / évolutions :** identifier les contenus qui font l'objet du plus de contestations, ou à l'inverse qui sont rapidement « abandonnés ».
- **Mesurer l'efficacité de la modération :** existe-t-il une différence significative entre, par exemple, le taux d'erreur annoncé par la plateforme et le taux d'erreur découlant du rapport des organes de résolution extrajudiciaires ? Ces chiffres sont-ils comparables ? Existe-t-il une « bonne métrique » sur la qualité de la modération ? L'objectif du tableau est d'avoir autant de facteurs possibles entre les mains.



Recommandations

Faire parler l'inter-transparence

Harmoniser, centraliser, standardiser : faire que les périodes, méthodes et catégories telles que notamment énoncées dans les modèles de rapports de transparence soient utilisées communément par l'ensemble des parties prenantes.

Extraire la valeur de l'inter-transparence : création d'une plateforme centralisée et communautaire accessible à toutes les parties prenantes pour créer de la valeur par la comparaison de toutes les sources d'information.

