

[FR version below]

## DSA Ad Repository Data Sprint

### Context

The Digital Services Act (DSA) was formally adopted in July 2022 and came into action on 17 February 2024. Among the various transparency measures introduced by the DSA, Very Large Online Platforms (VLOPs) have been obligated to create and maintain a Public Ad Repository that includes the content of each paid advertisement (text and image), information about the advertiser, typically including their identity, data regarding the target audience (demographic or interest-based parameters). The goal is to provide transparency that enables independent scrutiny and research regarding the influence of advertisements, particularly those that could impact democratic processes, public health, or societal debates. These repositories must be easily accessible for users, researchers, and regulators to scrutinize the nature and scope of advertising activities on the platform.

Nearly one year after the DSA has entered into force, the time has come to reflect on the real use of these ad repositories. The Open Institute for Digital Transformations, created as a part of [TIERED](#), proposes to do this via a data sprint tailored to enhance collaboration and knowledge sharing among academic scholars, regulators, journalists and activists. This collective experiment will allow these communities to explore various risks linked to online advertisements: misinformation, undeclared political discourse that may interfere with election processes, unethical corporate strategies, health-related campaigns, etc. Most VLOPS have established a public API which can be accessed by researchers, journalists or activists: among others, Facebook has extended its [Ad Library](#) and Tiktok is offering an API access to its [Commercial Content API](#), while Google has maintained its [Google Ads Transparency Center](#) allowing access to information about content that was moderated by the search engine.

### Format

We invite a wide array of experts from regulatory agencies and civil society organizations, data journalists, data scientists, social media researchers, legal scholars, sociologists, economists, computer scientists to work together with Sciences Po students in a Data Sprint to explore the possibilities offered by these repositories. During the sprint, teams consisting of students, engineers, and researchers from Sciences Po and interested partners will engage in an intensive program from February to April.

The “sprint” approach is aimed at preventing participants from becoming mired in exhaustive analysis, allowing them to focus on innovation and practical solutions. Teams will be guided by experts and engineers in using appropriate data sources, and data analysis techniques to answer a specific research question. The overarching

goal is to bolster the collaboration between research bodies, civil society and regulatory agencies, thereby creating an optimal environment for collaboratively developing potential policy recommendations for the successful rollout of the DSA.

Around five to ten teams will participate in three structured collaborative work sessions. The teams will be composed of the following members:

- Experts from academia, regulatory organizations, civil society, etc. with a familiarity with the substantive questions posed by the DSA / Ad repository
- Engineers / data scientists, who will be able to collect and analyze the data
- Students who are curious about the sprint, and who have good analytical skills and a basic knowledge of computational approaches"

Group leaders will be identified and selected in advance based on their interest and experience with the ad repository data. Masters' level students from Sciences Po will be selected, alongside the team leaders, based on their interest, skills, and motivation.

The sprint will be split into three phases, designed with distinct objectives and will include various group activities that will take place between February and April. Before the start of the first phase, group leaders are being identified and pre-work the data to determine feasibility of certain research questions. Teams will be formed during the first phase.

Phase 1 - *Project Design* (1 full day, **24 February**): Presentation of the objectives of the Data Sprint. The morning session will feature a certain number of keynote speakers: presentation of the DSA, specifically Article 39, and its challenges by Sciences Po researchers (ex. Beatriz Botero Arcila) and a regulator (ex. Arcom). Team Leaders will pitch their research questions and students will choose which group they want to join. We will make sure all teams will gather complementary expertise. In the afternoon, all groups will meet to discuss shared tasks and challenges as well as to establish guidelines to continue working remotely in their groups (selecting/acquiring the relevant dataset and skills, reviewing the literature, etc.).

Phase 2 - *Data-intensive Sprint* (3 afternoon sessions, **25, 26, 27 February**): In this stage, teams will gather to work on their dataset, addressing both informational and technical issues that arise during their analysis. Guidelines will be provided to structure the progress of each team and facilitate knowledge and skill transfer.

Phase 3 - *Results and Recommendations* (1 afternoon, **15 April**): The final phase involves teams presenting their results to an audience composed of researchers, policymakers, and regulators. Concluding remarks about collective recommendations for the ad repository reports will be produced, possibly through a collective policy brief.

Between each phase, groups are expected to work together virtually.

### **Participate in the data sprint**

If you are interested in participating in the data sprint, either as a researcher, student, engineer, regulator, or representative of civil society, we invite you to fill out [this form](#).

The deadline for submissions is **30 January**.

For any questions, please contact: [carly.hafner@sciencespo.fr](mailto:carly.hafner@sciencespo.fr).

---

## **Data sprint pour les registres publicitaire du *Digital Services Act***

### **Contexte**

Le *Digital Services Act* (DSA - Règlement sur les services numériques) a été officiellement adopté en juillet 2022 et est entré en vigueur le 17 février 2024. Parmi les diverses mesures de transparence introduites par le DSA, les très grandes plateformes en ligne (VLOP) ont été obligées de créer et de maintenir un registre publicitaire public qui comprend le contenu de chaque publicité payante (texte et image), des informations sur l'identité de l'annonceur, des données concernant le public cible (informations démographiques ou basée sur leurs intérêts). L'objectif est d'assurer une transparence qui permette un examen indépendant et des recherches sur l'influence des publicités, en particulier celles qui pourraient avoir un impact sur les processus démocratiques, la santé publique ou les débats de société. Ces registres doivent être facilement accessibles aux utilisateurs, aux chercheurs et aux régulateurs afin qu'ils puissent examiner la nature et la portée des activités publicitaires sur la plateforme.

Près d'un an après l'entrée en vigueur du DSA, le moment est venu de faire le point sur l'utilité de ces registres publicitaires. L'Institut libre des transformations

numériques, créé et soutenu par le projet [TIERED](#), organise un *data sprint* pour renforcer la collaboration et le partage des connaissances entre les universitaires, les régulateurs, les journalistes et les activistes. Cette expérience collective permettra à ces communautés d'explorer divers risques liés aux publicités en ligne : désinformation, discours politique non déclarés susceptible d'interférer avec les processus électoraux, stratégies d'entreprise contraires à l'éthique, etc. La plupart des VLOPs ont mis en place une API publique à laquelle peuvent accéder les chercheurs, les journalistes ou les activistes : Facebook a étendu sa [bibliothèque de publicités](#) et Tiktok offre un accès à son [API de contenu commercial](#), Snapchat propose des [bases de données](#) consolidés sur la publicité politique, tandis que Google a maintenu son [Centre de transparence des publicités Google](#), qui permet d'accéder à des informations sur le contenu modéré par le moteur de recherche.

## Format

Nous invitons un large éventail d'experts issus des mondes de la régulation, de la société civile, du journalisme, de la recherche (droit, économie des plateformes, sociologie et économie, informatique) à travailler avec des étudiants de Sciences Po dans le cadre d'un data sprint pour prendre part à cette initiative et explorer les possibilités offertes par ces registres. Pendant le data sprint, des équipes aux compétences mixtes s'engageront dans un programme intensif entre février et avril

L'approche « sprint » vise à développer des solutions pratiques et innovantes sur des sujets précis déjà pré-définis par des « chefs d'équipe ». Les équipes seront guidées par des experts et des ingénieurs dans l'utilisation de sources de données appropriées et de techniques d'analyse de données pour répondre à une question de recherche spécifique. L'objectif principal est de renforcer la collaboration entre les organismes de recherche, la société civile et les agences de régulation, créant ainsi un environnement optimal pour développer ensemble des recommandations politiques potentielles en vue d'une mise en œuvre réussie du DSA.

Cinq à dix équipes environ participeront à trois sessions de travail collaboratif structurées. Les équipes seront composées des membres suivants :

- Des experts du monde universitaire, des organismes de réglementation, de la société civile, etc. connaissant bien les questions de fond posées par le DSA / les registres publicitaires.
- Ingénieurs / scientifiques des données, qui seront en mesure de collecter et d'analyser les données

- Des étudiants curieux de ces données, avec de bonnes capacités analytiques et ayant une connaissance de base des approches computationnelles

Les chefs d'équipe seront identifiés en amont en fonction de leur intérêt et de leur expérience de travail pour ces données. Des étudiants de niveau master de Sciences Po seront sélectionnés, en concertation avec les chefs d'équipe, sur la base de leur intérêt pour ces données, de leur motivation et de leurs compétences analytiques.

Le sprint sera divisé en trois phases, conçues avec des objectifs distincts, et comprendra diverses activités de groupe qui se dérouleront entre février et avril. Avant le début de la première phase, les chefs de groupe sont identifiés et travaillent sur les données pour déterminer la faisabilité de certaines questions de recherche. Les équipes seront formées au cours de la première phase.

Phase 1 - *Conception du projet* (1 journée complète, **24 février**) : Présentation des objectifs du Data Sprint. La session du matin comprendra un certain nombre d'intervenants : présentation de du DSA, en particulier de l'article 39, et de ses défis par des chercheurs de Sciences Po (ex. Beatriz Botero Arcila) et un régulateur (ex. Arcom). Les chefs d'équipe présenteront leurs questions de recherche et les étudiants choisiront le groupe qu'ils souhaitent rejoindre. Nous veillerons à ce que toutes les équipes réunissent des compétences complémentaires. Dans l'après-midi, tous les groupes se réuniront pour discuter des tâches et des défis communs ainsi que pour établir des lignes directrices afin de continuer à travailler à distance dans leurs groupes (sélection/acquisition de l'ensemble de données et des compétences pertinentes, examen de la littérature, etc.)

Phase 2 - *Sprint* (3 sessions l'après-midi, **25, 26, 27 février**): Au cours de cette phase, les équipes se réuniront pour travailler sur leur ensemble de données, en abordant les questions informatives et techniques qui se posent au cours de leur analyse. Des lignes directrices seront fournies pour structurer les progrès de chaque équipe et faciliter le transfert de connaissances et de compétences.

Phase 3 - *Résultats et recommandations* (1 après-midi, **15 avril**): Lors de la phase finale, les équipes présenteront leurs résultats à un public composé de chercheurs, de décideurs politiques et de régulateurs. Des remarques finales sur les recommandations collectives pour les rapports du registre ad hoc seront produites, éventuellement par le biais d'une note d'information collective.

Entre chaque phase, les groupes sont censés travailler ensemble virtuellement.

Veillez noter que les activités du sprint se dérouleront **en anglais**.

### **Participer au data sprint**

Si vous souhaitez participer au data sprint, que ce soit en tant que chercheur, étudiant, ingénieur, régulateur ou représentant de la société civile, nous vous invitons à remplir [ce formulaire](#).

La date limite de soumission est fixée au **30 janvier**.

Pour toute question, veuillez contacter : [carly.hafner@sciencespo.fr](mailto:carly.hafner@sciencespo.fr).