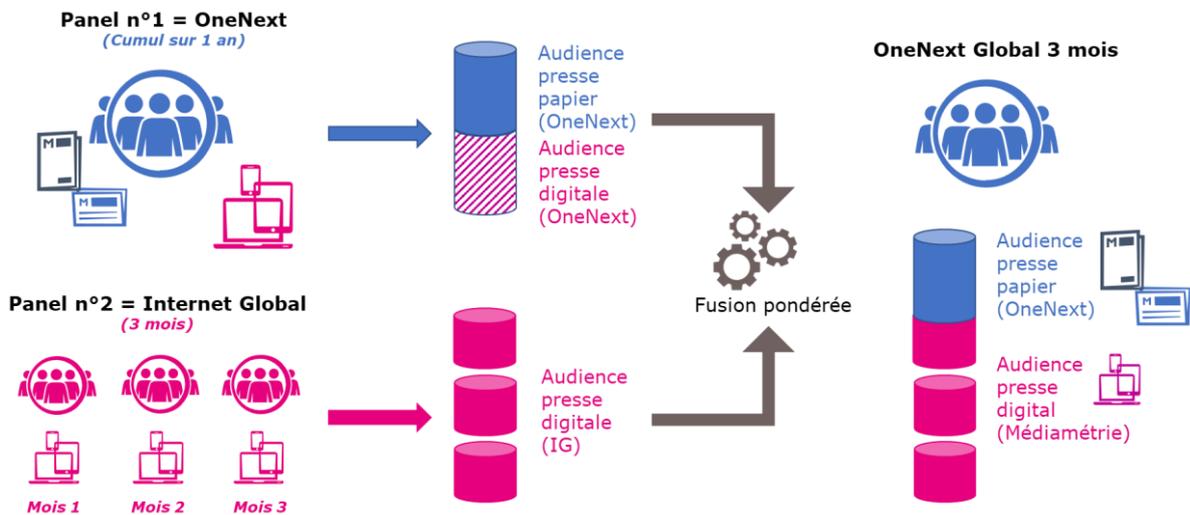


METHODOLOGIE

A partir de la vague OneNext Global S1 2022, Médiamétrie propose de cumuler en amont du process de fusion les 3 mois Internet Global (noté IG dans la suite du document) à considérer et de réaliser la fusion à l'identique de ce qui était fait sur les vagues précédentes en remplaçant le mois IG considéré par le cumul 3 mois IG. Le cumul est constitué par simple juxtaposition des 3 bases individuelles mensuelles, le poids des individus étant divisés par 3 afin de conserver les tailles des populations représentées. Le cumul constitué est ainsi représentatif du surf réalisé sur un mois moyen au sein de la période de 3 mois considérée.

SCHEMA D'ENSEMBLE DU DISPOSITIF



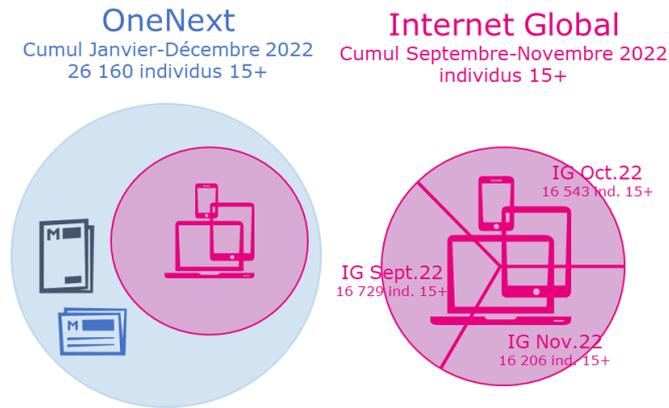
EXEMPLE DE LA PRODUCTION ONENEXT GLOBAL S1 2023

Donnée OneNext

- Cumul janvier-décembre 2022

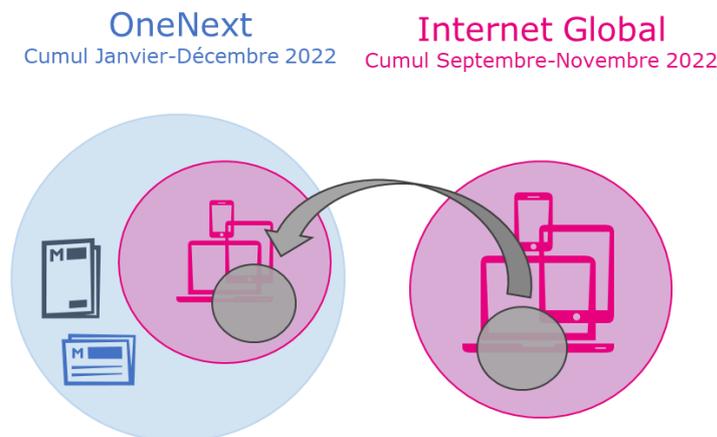
Données Internet Global considérées

- Vagues IG de septembre, octobre et novembre 2022



Univers concerné par la fusion

L'objectif de la fusion est de transférer les audiences des marques de Presse depuis le dispositif IG vers le dispositif OneNext. Il s'agit donc de transférer un sous-ensemble du dispositif IG vers un sous-ensemble du dispositif OneNext comme schématisé ci-dessous.



Pourquoi uniquement l'univers Presse et non l'ensemble des accédants à Internet ?

- 1 - il s'agit du périmètre à transférer vers OneNext.
- 2 - le transfert s'effectue par fusion et dépend donc de notre capacité à établir une ressemblance entre les individus OneNext et IG selon leur comportement de surf. Côté OneNext, seuls les comportements digitaux des marques de Presse sont renseignés dans

2

le questionnaire. Seul le surf des actifs Presse permet d'établir une similitude entre les individus des deux sources.

Comment sont définis les univers Presse sur chacune des deux sources ?

Un fichier « référentiel » permet de définir ces deux univers. Côté OneNext sont listées toutes les versions numériques associées à l'univers Presse digital et côté IG sont listés tous les éléments de la nomenclature associés à l'univers Presse digital.

Un individu OneNext sera dit *actif Presse* s'il a déclaré avoir visité au moins une version numérique du fichier référentiel.

Un individu IG sera dit *actif Presse* si du surf est remonté pendant le moins de mesure sur au moins une entité de la nomenclature du fichier référentiel

Méthode de fusion

Fusion par transport de poids

Afin de permettre une parfaite cohérence entre les niveaux d'audience Print (resp. digitales) mesurés sur OneNext (resp. IG) et ces mêmes niveaux après fusion avec le 3 dispositifs de référence des audiences digitales, Médiamétrie utilise une technique de fusion sous contrainte des niveaux Print et digitales.

Pour respecter les niveaux d'audience Print (resp. digitales) issus de OneNext (resp. IG), il suffit de conserver les poids associés aux individus de OneNext (resp. IG).

Sous la condition d'égalité des populations (ie effectifs pondérés) des actifs Presse entre OneNext et IG, on garantit de conserver après fusion les résultats digitaux (en milliers) sur cette population. En d'autres termes, si les poids des individus actifs Presse de OneNext sont cohérents (ie de même somme) avec ceux des individus actifs Presse d'IG, les audiences digitales (pour toutes les marques de Presse) sont assurées d'être conservées sur la cible 15 ans et plus.

Médiamétrie fournit à SocioDM le théorique 15 ans et plus actifs Presse issu de la vague IG considérée, permettant ainsi d'assurer la conservation des niveaux d'audience digitales sur la cible 15 ans et plus.

Afin de permettre la cohérence sur un plus grand nombre de cibles couramment utilisées sur le marché de la Presse, Médiamétrie décline cette approche sur une partition de la population des actifs Presse en 26 strates sociodémographiques croisant les critères de Sexe, Age et CSP de l'individu. Ainsi, l'algorithme de fusion précédemment décrit sur la population des Actifs Presse est décliné sur les 26 strates (on réalise donc 26 sous-fusions). On assure ainsi une conservation des résultats digitaux (pour toutes les marques de Presse) sur toutes les cibles combinaisons de ces 26 strates, sous réserve comme précédemment de l'égalité des populations (pondérées) sur chacune de ces 26 strates entre OneNext et IG. Médiamétrie transmet à SocioDM les structures des actifs Presse sur chacune des 26 strates, calculées à partir du cumul de 3 vagues IG (septembre, octobre et novembre 2021 sur la production OneNext Global S1 2022). La structure transmise est donc une structure moyennée sur 3 mois IG.

Les strates doivent être disjointes et de taille suffisante pour permettre au redressement de SocioDM de converger vers les théoriques fournis par Médiamétrie. Bien entendu, Médiamétrie transmet à SocioDM les théoriques associés.

Liste des strates sur lesquels sont effectuées les sous-fusions

strate	Sexe	Age_5	CSP_3
1	Homme	15-24 ans	CSPI+
2	Homme	15-24 ans	CSPI-
3	Homme	15-24 ans	Inactifs
4	Homme	25-34 ans	CSPI+
5	Homme	25-34 ans	CSPI-
6	Homme	25-34 ans	Inactifs
7	Homme	35-49 ans	CSPI+
8	Homme	35-49 ans	CSPI-Inactifs
9	Homme	50-64 ans	CSPI+
10	Homme	50-64 ans	CSPI-
11	Homme	50-64 ans	Inactifs
12	Homme	65 ans et +	CSPI+
13	Homme	65 ans et +	CSPI-Inactifs
14	Femme	15-24 ans	CSPI+
15	Femme	15-24 ans	CSPI-
16	Femme	15-24 ans	Inactifs
17	Femme	25-34 ans	CSPI+
18	Femme	25-34 ans	CSPI-
19	Femme	25-34 ans	Inactifs
20	Femme	35-49 ans	CSPI+
21	Femme	35-49 ans	CSPI-Inactifs
22	Femme	50-64 ans	CSPI+
23	Femme	50-64 ans	CSPI-
24	Femme	50-64 ans	Inactifs
25	Femme	65 ans et +	CSPI+
26	Femme	65 ans et +	CSPI-Inactifs

Choix de la distance

Médiamétrie a réalisé des tests de distance à utiliser dans la fusion. Nous avons testé une distance de Jaccard entre receveurs et donneurs car elle est adaptée aux événements très occasionnelles que sont les lectures d'un titre au regard de la population de l'étude. Cette distance ne prend pas en compte les similitudes sur les non-lectures et donne plus d'importance aux titres en commun lus par les individus. Les tests ayant été positif (amélioration du nombre de titres en commun entre les receveurs et les donneurs), nous avons reconduit l'utilisation de cette distance sur OneNext.

Distance de Jaccard

La *distance de Jaccard* entre deux individus i et j est définie par le complément à 1 de l'indice de similarité (indice de Jaccard) calculé à partir des lectures des deux individus :

$$d_{Jaccard}(i, j) = 1 - indice_{Jaccard}(i, j)$$

$$\text{avec } indice_{Jaccard}(i, j) = \frac{\text{nombre* de titres lus à la fois par } i \text{ et par } j}{\text{nombre* de titres lus par } i \text{ ou par } j}$$

A noter qu'à partir de la vague OneNext Global S1 2023, la façon de considérer la distance de Jaccard a légèrement évolué dans sa façon de traduire les lectures des titres selon le device (Ordinateurs, Mobile et Tablette). En effet, alors que sur les vagues antérieures, la distance retenue entre deux individus correspondait à la moyenne des distances de Jaccard calculées séparément sur chacun des devices, il a été décidé à partir de la vague One Next S1 2023 de calculer la distance indépendamment du device mais en considérant comme titre le croisement du titre et du device. Ainsi, un titre décliné sur 2 device est considéré au sein de la distance comme deux titres distincts. Il a été montré par Médiamétrie que cette distance améliorerait les indicateurs de contrôle considérés pour évaluer la qualité de la fusion.

Comme recommandé par ailleurs par le CESP lors des travaux sur ONE Global, nous avons conservé la pondération de chaque titre par l'inverse de son audience de sorte que la proximité entre deux individus lisant un même petit titre soit privilégiée. Les *nombre* de titres lus* définissant l'indice de Jaccard sont donc des nombres pondérés selon l'audience des titres (l'inverse de l'audience)

L'expression de la distance est donc la suivante :

$$d_{Jaccard}(i, j) = 1 - \frac{\sum_{d \in \{Ordi, Mob, Tab\}} \sum_{k=1}^N \frac{1}{\log(\text{Audience du titre}_k \text{ sur } d)^2} \cdot 1_{\{i \text{ et } j \text{ lisent le titre}_k \text{ sur } d\}}}{\sum_{d \in \{Ordi, Mob, Tab\}} \sum_{k=1}^N \frac{1}{\log(\text{Audience du titre}_k \text{ sur } d)^2} \cdot 1_{\{i \text{ ou } j \text{ lisent le titre}_k \text{ sur } d\}}}$$

La distance de Jaccard est comprise entre 0 et 1

Une distance nulle signifie que les deux individus lisent exactement les mêmes titres sur les mêmes devices alors qu'à l'inverse, une distance égale à 1 signifiera qu'ils ne lisent aucun titre en commun.

Cette distance a également l'avantage d'être plus facilement interprétable puisque directement lié au pourcentage de titres lus en commun entre deux individus parmi les titres lus par l'un ou par l'autre. Une distance de 0.1 entre deux individus signifie que 90% des titres lus (chaque titre étant pondéré selon son audience pour rappel) par l'un ou par l'autre sont lus par les deux.

Comme évoqué précédemment, la distance de Jaccard nécessite de définir pour chaque individu la liste de ces titres lus. A noter qu'il s'agit ici des versions digitales des titres lus.

Les individus à considérer venant de deux sources différentes, il est nécessaire de préciser la définition de la lecture selon le dispositif associé à l'individu.

Pour les individus issus du dispositif OneNext, on considère qu'un individu a lu un titre (digital) sur un type de device (Ordinateur, Mobile ou Tablette) s'il a déclaré le titre sur le device dans le questionnaire relatif aux versions numériques. A noter que comme dans le cadre de ONE Global, on dispose bien sûr de l'information de déclaration de lecture sur

les 30 derniers jours mais aussi la distinction entre une lecture plus de 30 jours et un « pas lu 12 mois ». Nous avons donc pu adapter la distance de Jaccard pour qu'elle considère une certaine proximité (via un facteur de pondération fixé à 0.5) entre un individu OneNext qui a déclaré lire un titre il y a plus de 30 jours mais non « pas lu 12 mois » et un individu IG qui a lu ce titre sur le mois considéré par la fusion.

Liste des variables de pont considérées dans le calcul de la *distance de Jaccard*

Les variables de pont sont les informations permettant de calculer la distance de Jaccard entre les individus de OneNext et les individus IG. Nous avons abordé dans le paragraphe précédent la nature des données utilisées : les versions numériques pour les individus OneNext et le surf remonté dans IG sur le mois considéré pour les individus IG

Reste à préciser la façon dont nous identifions l'élément de nomenclature IG à associer à chaque version numérique d'un titre de Presse de OneNext et ainsi permettre de définir si deux individus des deux sources différentes lisent les mêmes titres et donc calculer la distance entre les deux individus.

Le fichier référentiel est le fichier répondant à ce besoin. Il met en relation les marques de Presse, les versions numériques associées et les éléments de nomenclature IG correspondants. Il est mis à jour à chaque nouvelle fusion et validé par l'ACPM.