

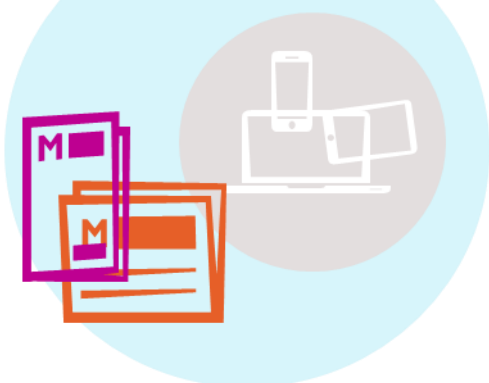
## METHODOLOGIE

### Donnée Internet Global considérée pour la donnée intermédiaire

- Vague Internet Global de Octobre 2019

### Schéma d'ensemble du dispositif

**One Next**  
Cumul Mars-Décembre 2019  
27 899 individus  
Ens 15+

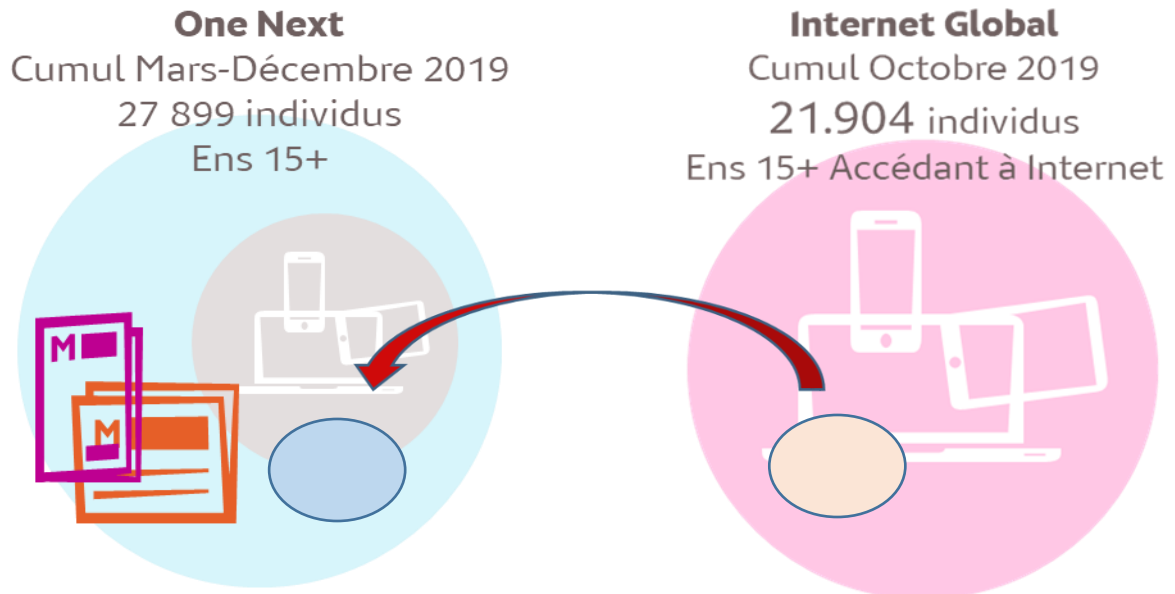


**Internet Global**  
Cumul Octobre 2019  
21.904 individus  
Ens 15+ Accédant à Internet



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

## Univers concerné par la fusion



### 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 Internet Global selon leur comportement de surf. Côté ONE, seuls les comportements digitaux des marques de Presse sont renseignés dans 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 Internet Global 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 Internet Global 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. Internet Global) 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. Internet Global), il suffit de conserver les poids associés aux individus de OneNext (resp. Internet Global).

Sous la condition d'égalité des populations (ie effectifs pondérés) des actifs Presse entre OneNext et Internet Global, 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'Internet Global, 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 Internet Global 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 Internet Global. Médiamétrie transmet à SocioDM les théoriques des actifs Presse sur chacune des 26 strates, calculés à partir de la vague Internet Global considérée (ici Octobre 2019).

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 :

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

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

Comme recommandé par le CESP lors des travaux sur ONE Global, nous avons pondéré 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)

La distance de Jaccard est comprise entre 0 et 1

Une distance nulle signifie que les deux individus lisent exactement les mêmes titres 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'autres 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 dans une première hypothèse (H0) 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 jamais. 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 ONE qui a déclaré lire un titre il y a plus de 30 jours mais non jamais et un individu Internet Global 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 ONE et les individus Internet Global. Nous avons abordé dans le paragraphe précédent la nature des données utilisées : les versions numériques pour les individus ONE et le surf remonté dans Internet Global sur le mois considéré pour les individus Internet Global

Reste à préciser la façon dont nous identifions l'élément de nomenclature Internet Global à associer à chaque version numérique d'un titre de Presse de ONE 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 Internet Global correspondants. Il est mis à jour à chaque nouvelle fusion et validé par l'ACPM.

### **Indicateurs de contrôle envisagés (liste non exhaustive en cours de construction)**

- Vérification de l'atteinte des théoriques des 26 strates sur les poids individuels transmis par SocioDM.
- - Vérification du respect des audiences post-fusion de chaque élément de nomenclature du fichier référentiel sur l'ensemble des 26 strates par rapport aux résultats Internet Global considérés.
- - Calculs des indicateurs de contrôle partagés avec le CESP lors des productions ONE Global (% de titres lus en commun, ...)
- - Répartition des actifs Presse selon le device et comparaison de cette répartition entre la donnée OneNext et la donnée Internet Global
- - Distribution des distances post-jumelage sur chacune des strates.
- - Représentation distances issues du jumelage par rapport à la distance min
- - Validation en sortie de jumelage que les poids des individus ONE et des individus Internet Global sont conservés (et que tous les individus ont été sélectionnés au moins une fois dans le jumelage) => On valide ainsi que l'algorithme d'optimisation a respecté les contraintes.
- - Comparaison des affinités digitales sur les 115 cibles (normalement nulles sur toutes les cibles issues de 26 strates) et espérons raisonnables sur les autres.