

Fusion des études de référence de la mesure d'audience Radio, TV, Internet et Presse

Aurélie Vanheuverzwyn

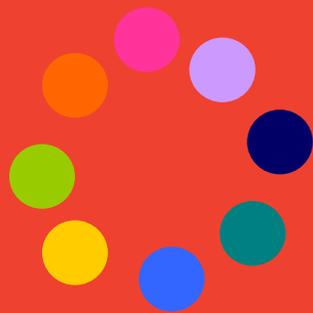
JMS, 25 mars 2009





1. Contexte et objectifs
2. Méthodologie retenue
3. Retour d'expérience après la livraison de la première vague de résultats

Contexte et objectifs





Le contexte

Evolution du mode de consommation média

- ▶ Des supports d'accès aux médias plus largement disponibles
- ▶ Des médias plus largement diffusés

Fragmentation des audiences

- ▶ En TV, avec l'apparition des chaînes de la TNT, l'offre restreinte est passée de 7 à 18 chaînes
- ▶ Sur Internet, l'offre est illimitée
- ▶ En Radio, développement des webradios et bientôt, arrivée de la radio numérique

La marque média se décline sur les différents supports

- ⇒ Besoin de visibilité sur sa couverture tous supports confondus



Le besoin

Demande initiale

▶ Apporter un premier éclairage sur les comportements plurimédia des français

En 2005, naissance de l'étude « Media In Life »

▶ Recueil, sur une journée, de l'ensemble des activités média et multimédia quart d'heure par quart d'heure auprès d'un échantillon de 5000 personnes

▶ Limites :

- Absence de profondeur temporelle
- Pas de marque

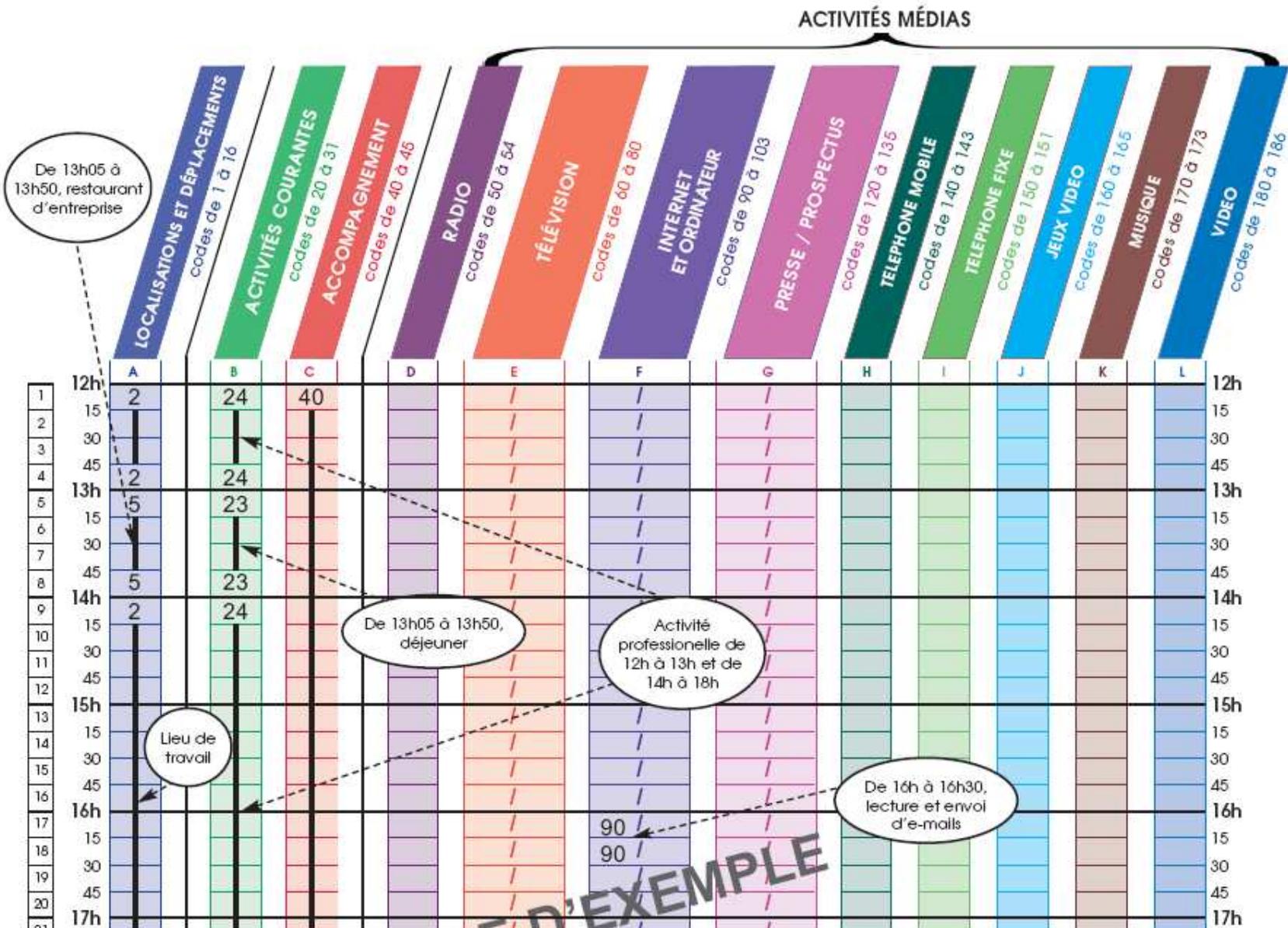
Première réponse permettant d'arbitrer entre les différents médias...

...mais difficilement exploitable dans une optique médiaplanning



Journal d'activités de l'enquête

Média In Life





L'étude Cross-Médias

Nouveau cahier des charges

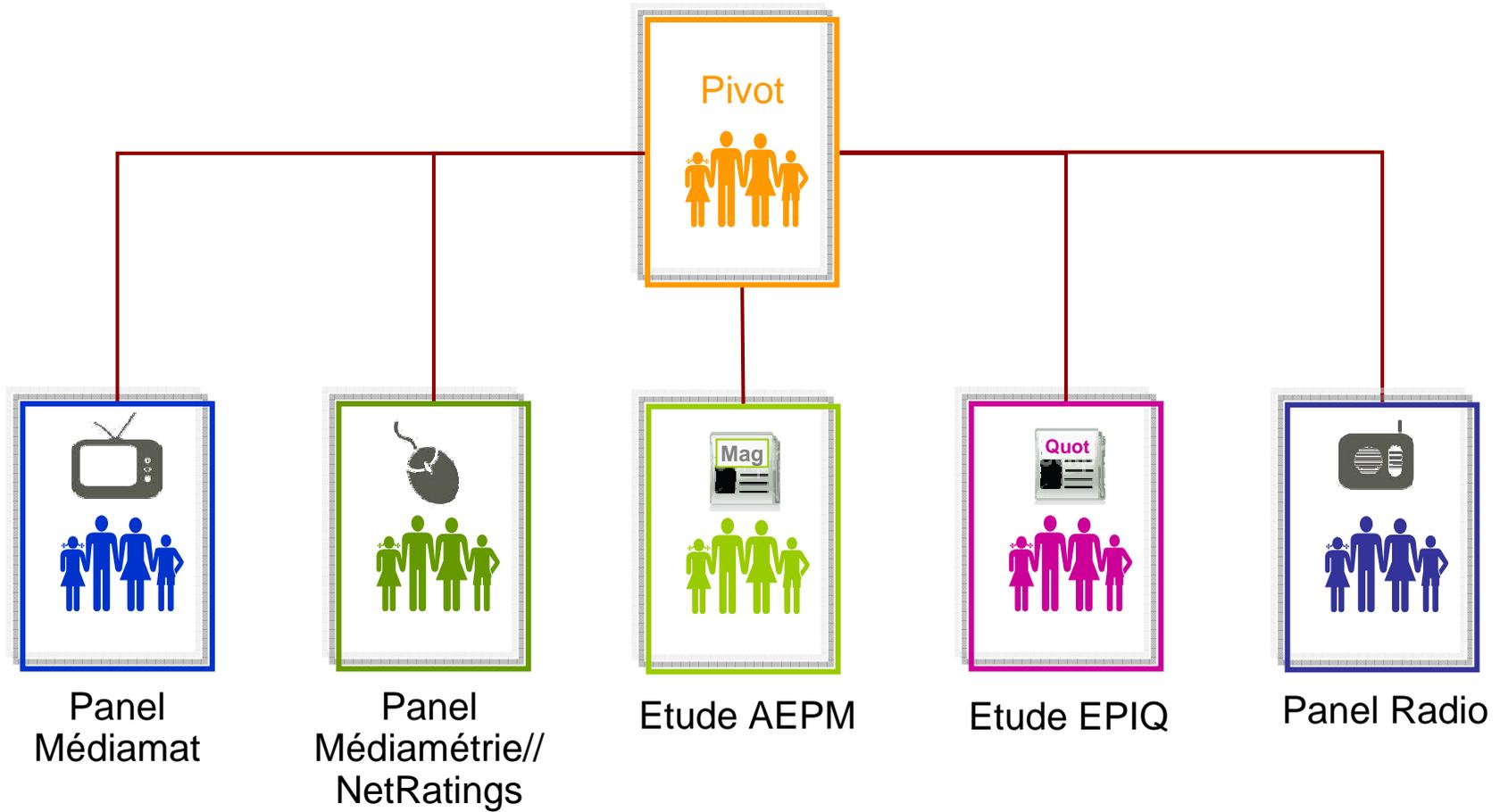
- ▶ Construire, dans un cadre économique réaliste, un dispositif s'appuyant les enquêtes de référence mono-média
- ▶ C'est-à-dire :
 - Mettre au point une étude permettant de disposer, sur plusieurs médias, de résultats de même nature que les enquêtes de référence
 - Assurer la cohérence des résultats de cette étude avec ceux des enquêtes de référence

La réponse apportée par Médiamétrie

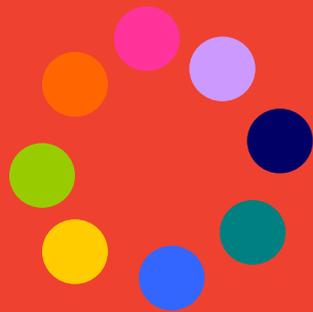
- ▶ Créer, par fusion statistique des enquêtes de référence mono-média autour d'une enquête « pivot », une base d'individus virtuels dont on observerait les comportements d'audience TV, Radio, Internet et Presse



Schématiquement



Méthodologie





Le concept de fusion

Principe

- ▶ Injecter au fichier dit « receveur » un certain nombre de variables présentes dans le fichier « donneur »
- ▶ Les variables communes aux deux fichiers sont les variables relais
- ▶ Les variables propres à chaque fichier sont les variables spécifiques

Deux grands types de méthodes

- ▶ Approche par modèle
 - Chaque variable spécifique est étudiée de manière indépendante
 - Limite : les corrélations observées dans le fichier donneur sont difficilement maintenues dans le fichier receveur
- ▶ Appariement
 - Détecter, pour chaque individu du fichier receveur, un ou plusieurs « sosies » au sein du fichier donneur

Déterminer les individus qui se ressemblent
et les assembler dans une base unique



Les grandes étapes

Etape 1 : sélection des variables relais

- ▶ Lister ou construire les variables communes aux deux fichiers
- ▶ Les hiérarchiser en fonction de leur pouvoir explicatif sur les variables spécifiques :
 - certaines ne seront pas retenues comme variables relais du fait de leur faible pouvoir explicatif
 - d'autres seront considérées comme prioritaires / obligatoires : ce sont les variables de contrôle
- ▶ Au sein de chaque strate, donneurs et receveurs seront appariés en fonction de leur similitude au regard des variables relais



Les grandes étapes

Etape 2 : choix de la fonction de distance

▶ Distance de Manhattan : $\sum_{i=1}^n |x_i - y_i|$

▶ Distance euclidienne : $\sqrt{\sum_{i=1}^n (x_i - y_i)^2}$

▶ Distance de Levenshtein :

- distance intégrant une fonction de coût permettant de pénaliser les différences en tenant compte de la probabilité d'apparition d'une valeur

▶ Distance de Jaccard :

- distance permettant de mesurer la proximité entre deux jeux de variables binaires



Les grandes étapes

Etape 3 : sélection des sosies

- ▶ Contrainte : l'ensemble des observations des deux bases de données doivent apparaître dans la base fusionnée

- ▶ Processus de détermination des sosies :
 1. On affecte à chaque receveur son donneur le plus proche
 - ⇒ Un même donneur peut être affecté plusieurs fois
 2. On affecte les donneurs non encore attribués aux receveurs qui leurs sont les plus proches
 - ⇒ A l'issue de ces deux étapes, donneurs et receveurs peuvent apparaître plusieurs fois
 3. On supprime les couples inutiles



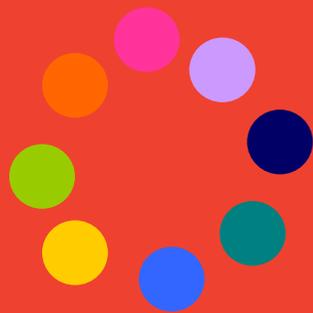
Validation

Etape 4 : validation de l'imputation

- ▶ Validation de l'appariement
 - Distribution des distances entre donneur et receveur
 - Nombre d'apparitions des donneurs et des receveurs dans la base finale

- ▶ Cohérence avec les bases d'origine
 - Moyennes / distributions marginales
 - Test de Wilcoxon de comparaison des rangs
 - Test de Mantel de comparaison des corrélations

Retour d'expérience après la livraison de la première vague de résultats





Application

Description des données étudiées

- ▶ Enquête pivot :
 - Issue du panel Radio de Médiamétrie
 - 5000 individus
 - Informations socio-démographiques et géographiques
 - Habitudes d'écoute TV par chaîne et par émission, fréquence de connexion Internet au global et sur quelques sites, habitudes de lecture de la presse



Application

Description des données étudiées

- ▶ Enquêtes sources :
 - Panel Médiamat pour la TV
 - Panel Médiamétrie//NetRatings pour Internet
 - Enquête AEPM pour la presse magazine
 - Enquête EPIQ pour la presse quotidienne
 - Panel Radio pour la Radio
- ⇒ Fusion « informatique » avec le pivot



Difficultés rencontrées

Modes de recueil ou questionnaires différents entre enquête pivot et enquêtes sources

- ▶ TV : déclaratif dans l'enquête pivot vs mesure automatique dans le panel Médiamat
- ▶ Internet : idem + perception du périmètre des sites
- ▶ Presse : questionnaire plurimédia vs questionnaire dédié

⇒ Les variables communes sont-elles vraiment comparables ?



Difficultés rencontrées

Plans de sondage et tailles d'échantillon différents d'une enquête à l'autre

- ▶ Rapport donneurs / receveurs différent d'une fusion à l'autre
- ▶ Pour une même fusion, le rapport est hétérogène car les strates n'ont pas forcément le même taux de sondage

⇒ Complexité de la validation de l'appariement !



Difficultés rencontrées

Chaque fusion est réalisée de manière indépendante

- ▶ A l'issue des fusions, un individu de l'enquête pivot s'est vu affecter par exemple 5 sosies pour la TV, 4 pour Internet, 1 pour la presse quotidienne et 2 pour la presse magazine
- ▶ On crée donc autant d'individus virtuels que de combinaisons possibles

⇒ Quels sont les effets de ce dépliage multiplicatif sur les corrélations ?



Difficultés rencontrées

Validation des résultats

- ▶ Possibilité de valider les résultats des fusions une à une
- ▶ Mais aucune source pour valider la pertinence des duplications obtenues entre supports de différents médias

⇒ Comment s'assurer de la transitivité de la relation ?



Malgré cela...

Retours clients positifs après deux mois d'exploitation

- ▶ Livrable en adéquation à leurs attentes
- ▶ Résultats jugés « cohérents »

De nouveaux besoins en découlent

- ▶ Evaluation de la précision des résultats de cette étude

Livraison de la prochaine vague de résultats en juin