

Mesure d'audience et données massives : mythes et réalités

Aurélie Vanheuverzwyn, Médiamétrie



Médiаметrie

9^{ème} Colloque Francophone sur les Sondages
Gatineau, le 13 octobre 2016

1

**Données massives : de quoi
parle-t-on ?**



« Big data : l'or noir du XXIe siècle »



De quoi parle-t-on ?



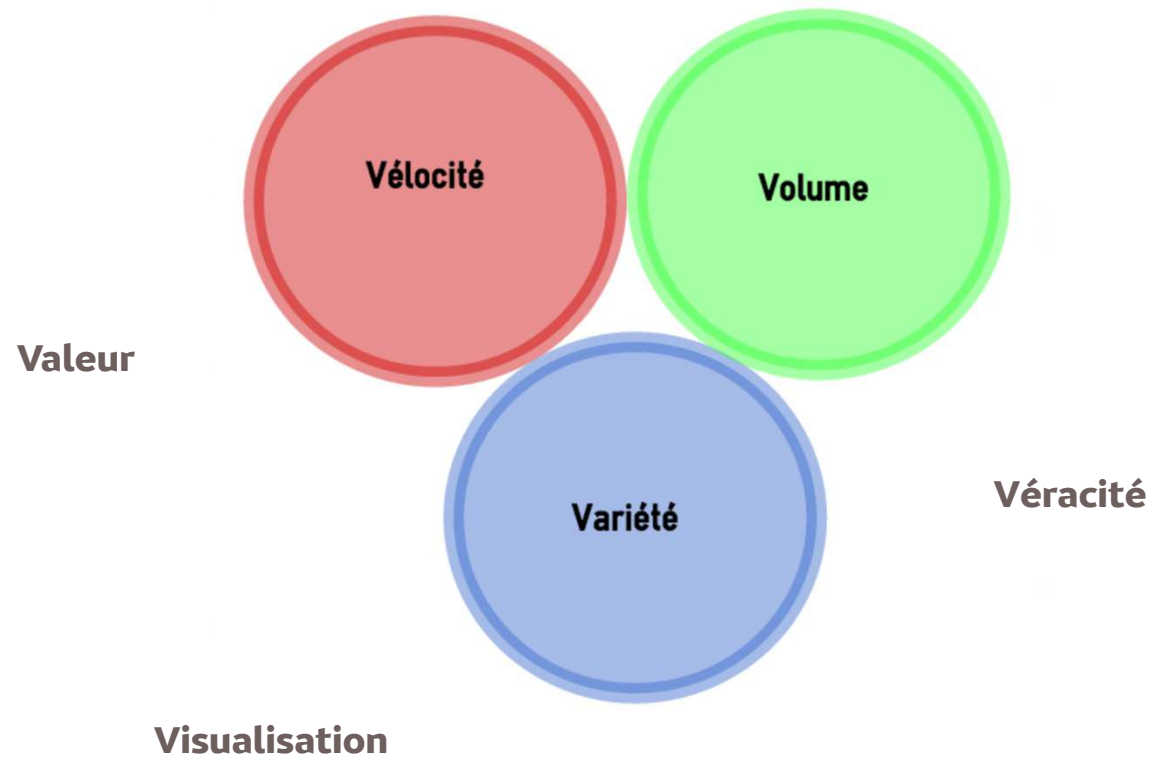
Les « données massives » désignent l'ensemble des données numériques produites par l'utilisation des nouvelles technologies :

- *données issues de capteurs,*
- *contenus publiés sur Internet (images, vidéos, sons, textes),*
- *transactions de commerce électronique,*
- *échanges sur les réseaux sociaux,*
- *données transmises par les objets connectés (téléphones mobiles, compteurs intelligents,...),*
- *données de géolocalisation,*
- *etc...*

De quoi parle-t-on ?



Des ensembles de données tellement volumineux qu'ils en deviennent difficiles à travailler avec des outils classiques de gestion de base de données ou de gestion de l'information.



De vrais enjeux et opportunités de recherche



Un thème de recherche et de développement en informatique :

- *De nouveaux modèles de représentation des données*
- *De nouvelles architectures de stockage*

Un vrai potentiel pour la recherche scientifique :

- *Traitement des données astronomiques, génétiques, météorologiques...*
- *Secteurs de la santé, de l'énergie, de l'habitat...*

Mais aussi des opportunités commerciales



Meilleure connaissance des clients

Ciblage en temps réel en fonction des probabilités d'achat

Automatisation des décisions

2

Les médias se numérisent et leur consommation également



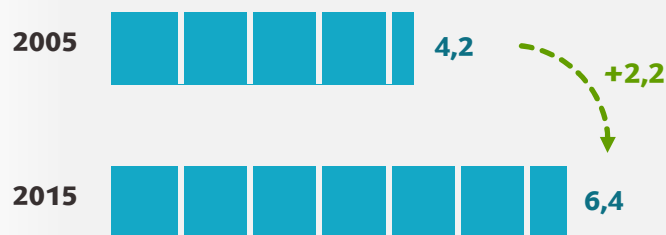
Des français très connectés



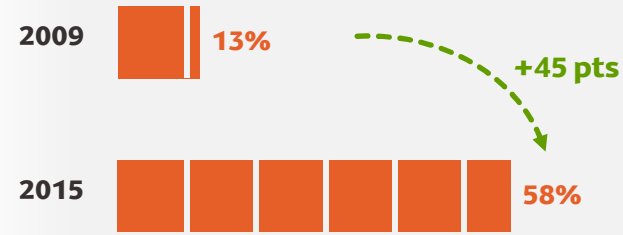
Depuis 10 ans nous assistons à une multiplication des écrans dans les foyers. Le développement des écrans mobiles permet d'accéder à des nouvelles offres et de nouveaux usages.



Nombre d'écrans par foyer



Taux d'équipés de smartphones



Nombre de contacts Médias et Multimédias sur une journée

Périmètre = TV, Radio, Internet, Presse, Cinéma, Jeux vidéo, Vidéo, Musique, Téléphone



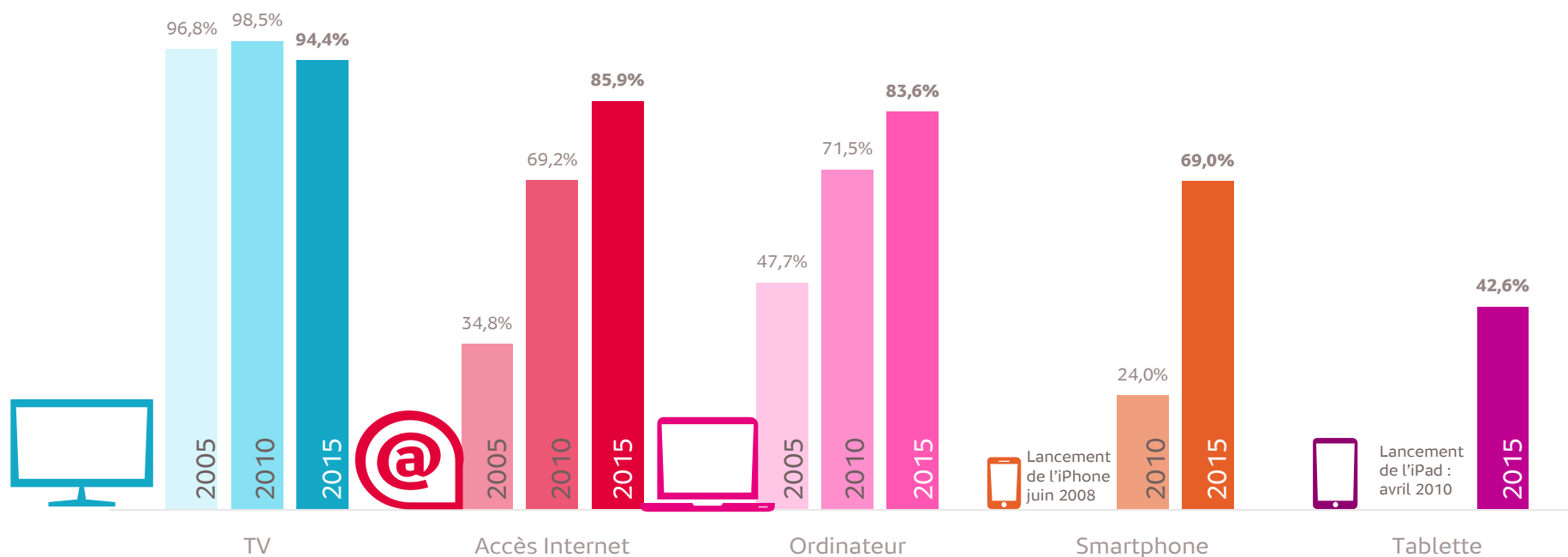
Source : Home Devices T2 2015 REM Médiamétrie/GFK T2 2005 Base Foyers / Média In Life 2014 -13 ans et +, Lundi Dimanche, 00H00-24H00, Avec ou sans act. courante, Toutes localisations, Avec ou sans accompagnement

Les équipements : la TV en retrait, internet en hausse



L'équipement en postes de télévision au sein des foyers français est en recul depuis 5 ans. Dans la même période, l'accès à internet dans les foyers français poursuit sa croissance. Parallèlement, les équipements mobiles (tablettes et smartphones) explosent au cours des 5 dernières années

Équipement Multimédia et accès Internet



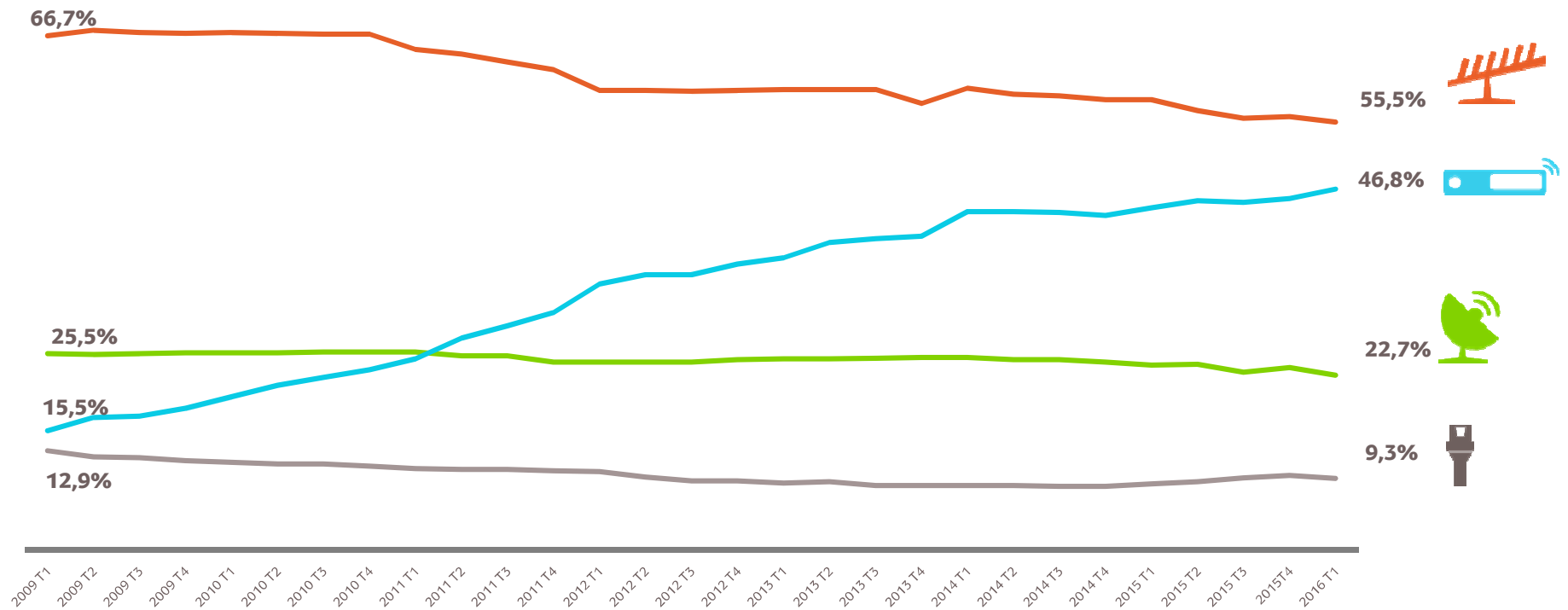
Source : TV / Accès Internet / Ordinateur / Tablette = REM en 2010, Home Devices en 2015 (équipement du foyer)
Smartphone = Webo Global, Internauts 15+ (équipement individu)

Une évolution de la réception TV par ADSL très importante au cours des 5 dernières années



Les modes de réception de la télévision, qui ont fortement évolué ces dernières années, offrent de nouvelles opportunités pour les téléspectateurs : le différé, la catch-up, un nombre de chaînes croissant...

Evolution des modes de réceptions TV en % de foyers équipés TV



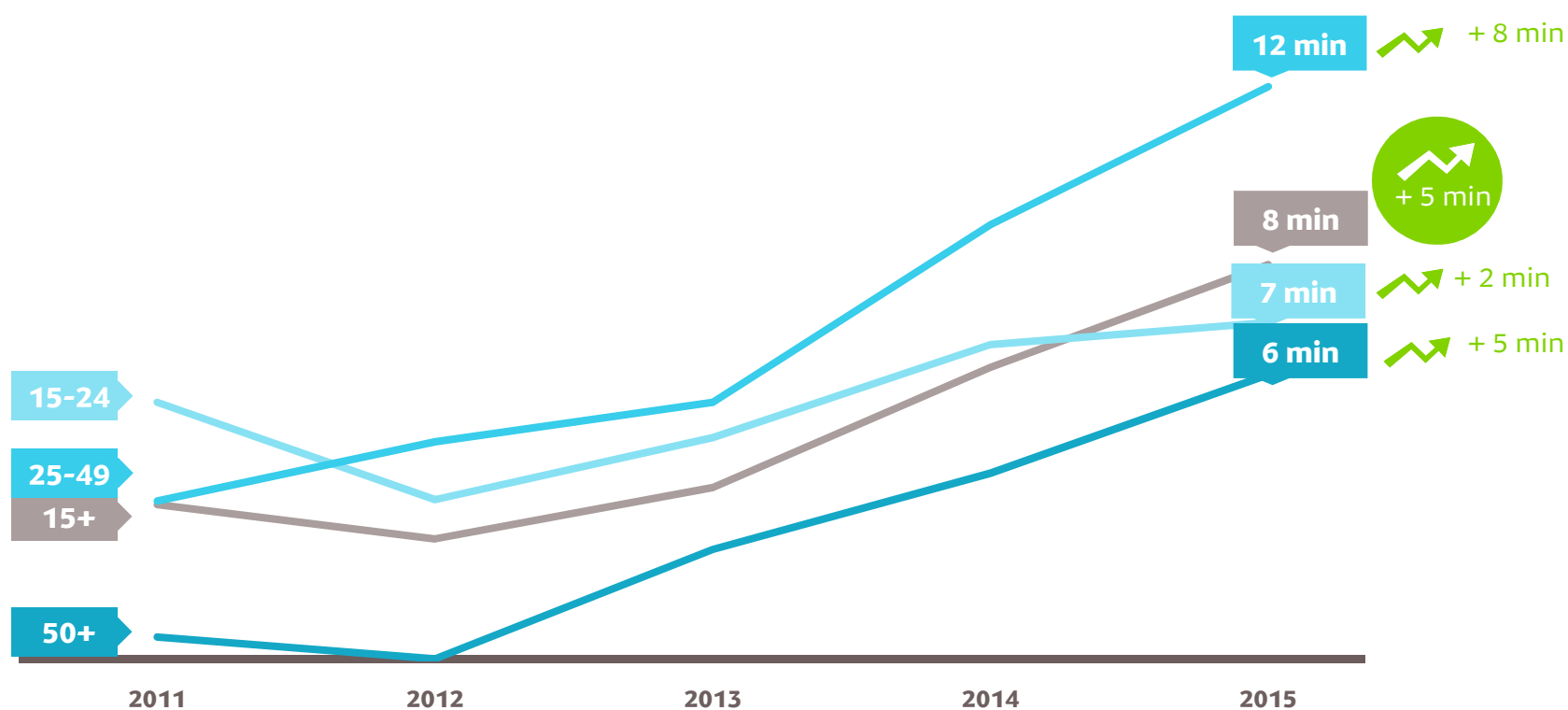
Source : REM Médiamétrie –GFK 2009-2013 Home Devices 2014-2016

Les nouveaux usages de la télévision – Une durée de consommation en forte progression



Quelle que soit la population, les nouvelles façons de regarder la télévision : en catch-up ou sur d'autres écrans, constituent une vraie alternative à la consommation de la télévision en live ou en différé sur le poste de télévision. La DEI de ces nouveaux usages a été multipliée par 3 en 4 ans.

Evolution de la durée d'écoute par individu des nouveaux usages TV : catch-up sur le téléviseur ou TV sur internet, 2011 vs 2015

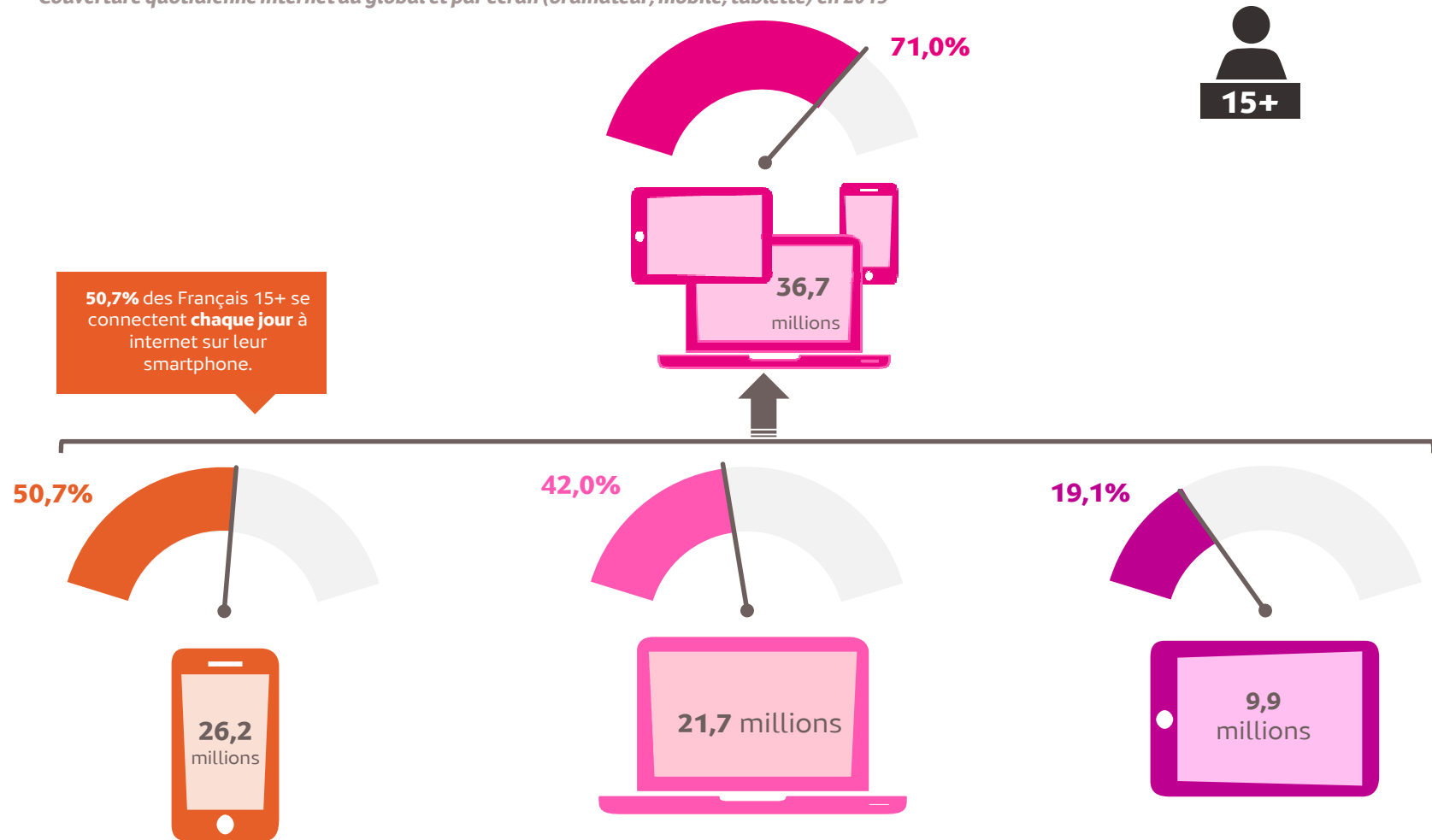


36,7 millions d'internautes de 15 ans et plus chaque jour



En 2015, **36,7 millions**, soit **71%** de Français âgés de 15 ans et plus se sont connectés à internet via un ordinateur et/ou un mobile et/ou une tablette sur une journée moyenne.

Couverture quotidienne internet au global et par écran (ordinateur, mobile, tablette) en 2015



Source : Médiamétrie et Mediametrie//NetRatings - Données France entière recalculées à partir des données des Panels Audience Internet Ordinateur, Audience Internet Mobile, Audience Internet Tablette et Audience Internet Global - France - moyenne 2015

L'achat d'espace publicitaire évolue en conséquence



De l'achat media à l'achat d'audience



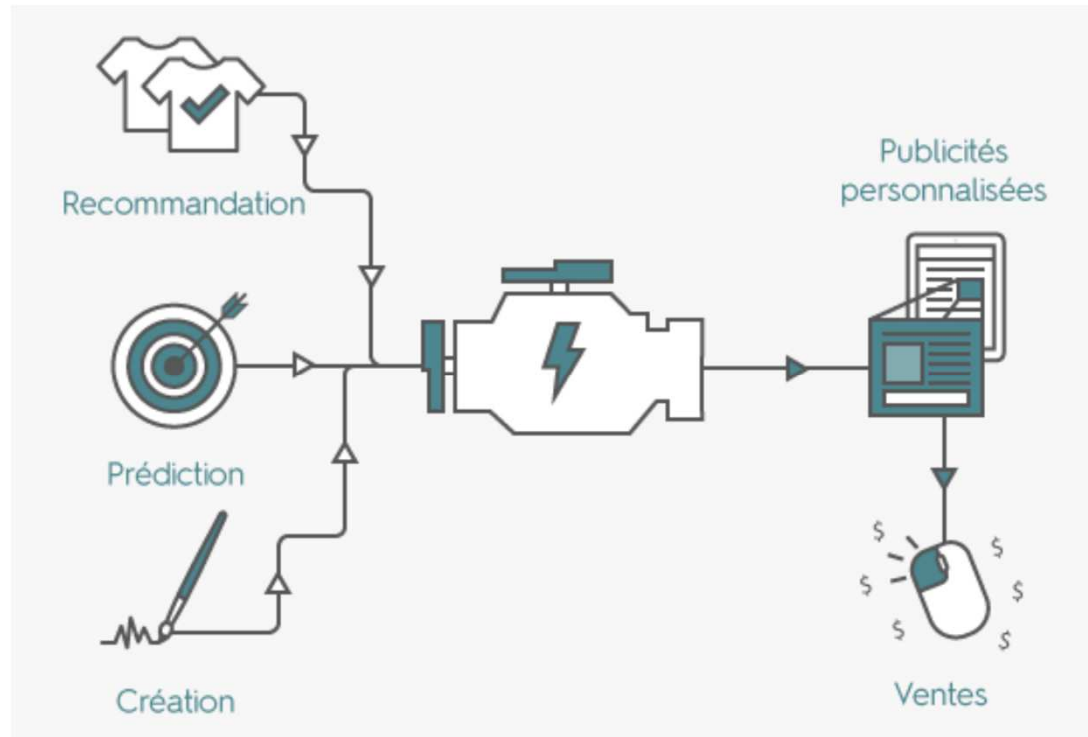
Les modèles d'achat d'espace publicitaire traditionnels reposaient sur la sélection du support le plus en affinité avec la cible recherchée par l'annonceur.

Aujourd'hui, les modèles de profilage permettent aux annonceurs de cibler directement l'individu recherché.

L'achat d'espace publicitaire évolue en conséquence



Exemple : Critéo



Ciblage des utilisateurs les plus susceptibles d'effectuer un achat

Personnalisation de la publicité délivrée à l'internaute

3

La fin des mesures d'audience ?



A quoi sert la mesure d'audience ?



L'audience représente l'ensemble des personnes touchées par un média ou support de communication. Il peut s'agir de téléspectateurs, d'auditeurs ou de visiteurs d'un site Internet.

La mesure d'audience répond à deux besoins principaux :

- Des besoins éditoriaux

Les résultats d'audience permettent de qualifier et/ou quantifier l'auditoire d'une émission, d'un site, d'identifier les principaux carrefours d'audience ainsi que les zones d'ombre et d'ajuster en conséquence la politique éditoriale.

- Des besoins publicitaires

La mesure d'audience est un outil d'arbitrage pour les annonceurs qui, dans le cadre de leurs campagnes publicitaires, sélectionnent les supports dont les critères d'audience correspondent le mieux à leurs objectifs de communication.

Comment fonctionne la mesure d'audience ?



Pour la TV et Internet :

1. *Recrutement d'un échantillon*
2. *Installation et/ou téléchargement d'un audimètre/logiciel de mesure*
3. *Suivi longitudinal de l'audience individuelle des panélistes*
4. *Traitement quotidien/mensuel des données*

Pour la Radio :

1. *Recrutement d'un échantillon*
2. *Passation d'un questionnaire sur l'audience au cours des dernières 24 heures*
3. *Traitement trimestriel des données*

La mesure d'audience de plus en plus critiquée



Les principales critiques :

L'offre de contenus est maintenant trop vaste pour être correctement mesurée auprès d'un échantillon.

Le déclaratif n'est pas fiable.

Les dispositifs de mesure sont trop contraignants pour que les échantillons soient « représentatifs ».

En synthèse :

- 1. Des échantillons trop petits**
- 2. Des biais de déclaration**
- 3. Des taux de réponse trop faibles**

Données massives = la réponse aux faiblesses des enquêtes par sondage

Les problèmes liés aux données massives



Quantité ne signifie pas nécessairement qualité

Les données numériques recensent généralement des machines, y compris des robots, et non des individus.

Accessible ne veut pas dire éthique

Les données numériques sont souvent collectées à l'insu des individus.

Le droit protégeant les données à caractère personnel apparaît aujourd'hui comme le pire ennemi des mégadonnées.

L'accès limité aux mégadonnées crée de nouvelles fractures numériques

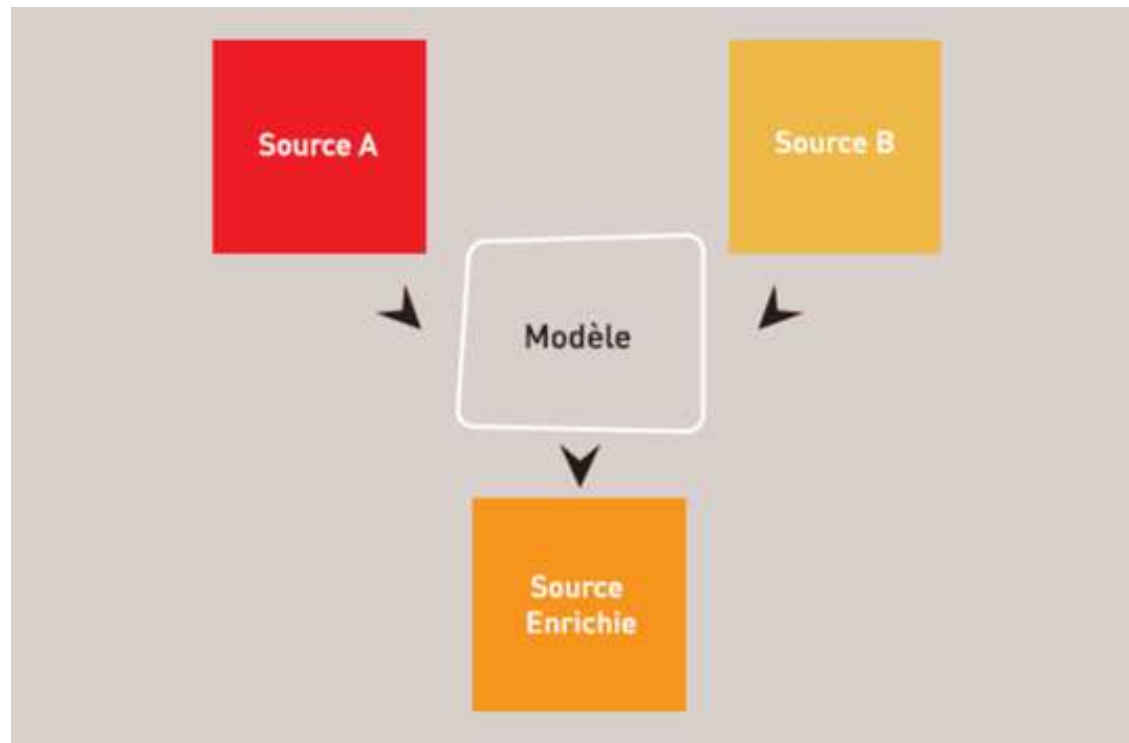
Collecter et analyser ses propres données demande des moyens financiers importants.

Les entreprises ayant cette capacité sont finalement peu nombreuses.

La fin des mesures parfaites et l'émergence des mesures hybrides



Un concept simple : utiliser toutes les sources d'informations disponibles pour produire des résultats plus riches, plus fiables...



Exemples de méthodes hybrides



La mesure d'audience d'Internet en France

Les données exhaustives viennent enrichir les données d'enquête

1. **Un panel d'individus dont l'activité sur Internet est mesurée à l'aide d'un logiciel installé sur leur(s) ordinateur(s), téléphone(s) mobile(s) et tablette(s)**
2. **Des comptages exhaustifs réalisés grâce à l'insertion de marqueurs sur l'ensemble des pages des sites Internet**



Calage du panel sur les résultats de la mesure exhaustive

Exemples de méthodes hybrides



La mesure hybride des chaînes thématiques

Les données d'enquête viennent enrichir les données exhaustives

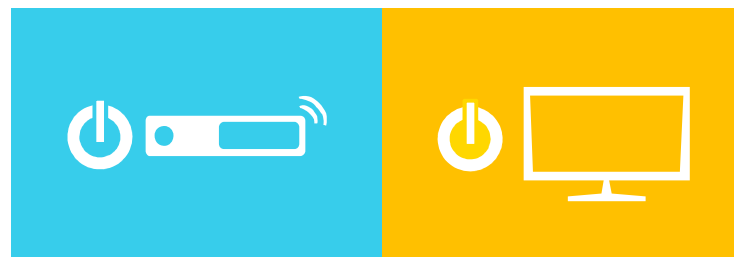
1. **Des logs (disponibles pour les décodeurs numériques ADSL, Câble et Satellite connectés à Internet) permettant de savoir l'heure à laquelle la consommation d'une chaîne ou d'un service a débuté**
2. **Un panel d'individus dont l'audience de la télévision est mesurée à l'aide d'un audimètre installé sur leur(s) téléviseur(s)**



Transformation de la donnée d'usage en donnée d'audience individuelle grâce à une modélisation en 3 étapes.



Etape 1 : Passage décodeur / téléviseur



Décodeur ON → TV ON

Pré-traitement des logs bruts de l'échantillon

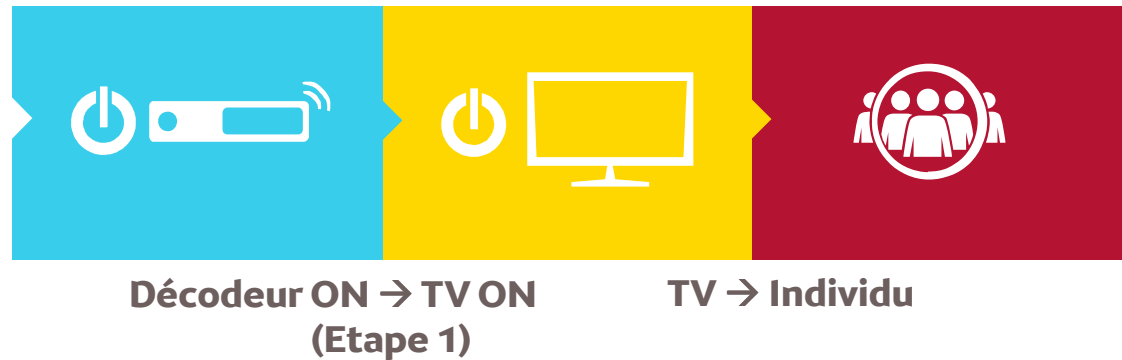
- Seuls les usages correspondant à de la consommation TV sont conservés => Suppression des logs « techniques », des usages VOD, services interactifs ...
- Constitution de tickets d'audience : heure de début, heure de fin, chaîne

Écrêtage

- Estimation et suppression des usages correspondant à du TV Off
- Paramètres de la fonction d'écrêtage estimés à partir de Médiamat



Etape 2 : Individualisation des audiences téléviseur



Une modélisation sociodémographique et comportementale

Objectif : Déterminer, à chaque instant, qui regarde la télévision

Pré-requis : Qualification sociodémographique de l'échantillon traité

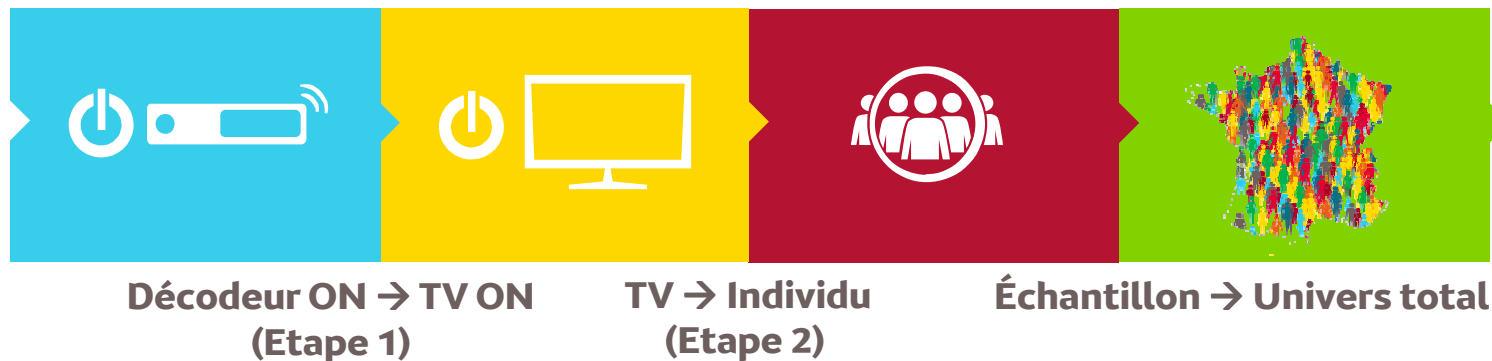
Paramètres du modèle

Estimés à partir des observations de Médiamat (sur le même univers)

Regroupement des chaînes par genre pour une modélisation plus robuste



Etape 3 : extrapolation à l'univers total



Redressement de l'échantillon

L'échantillon des abonnés voie de retour doit être représentatif de l'ensemble des abonnés sur des critères socio-démographiques et d'équipement

4

Conclusion et perspectives



La fin des mesures parfaites



Des taux de réponse en baisse constante, en particulier sur les populations les plus jeunes

Une confiance altérée dans les sondages



Nécessité de faire évoluer les méthodologies et à aller vers des systèmes d'observation de plus en plus complexes.

Quelques pistes de réflexion



Autoriser la non-réponse partielle : vers des questionnaires à la carte ?

Développer le multimode

Utiliser des informations exogènes nouvelles au niveau de l'échantillonnage

Au-delà des enjeux méthodologiques, des questions de société



Valeur de la donnée, partage, éthique, protection de la vie privée... Si les données massives apportent de nouvelles perspectives, elle soulève aussi des questions de société et de modèle auxquelles nous devons répondre.

Merci pour votre attention

