

Le Big Data pour connaître et engager ses utilisateurs

Breizh Data Day
19/10/2017



Sommaire

-  **1.** **So... what?**
-  **2.** **Outils et cadre juridique**
-  **3.** **Programme RETURN!**
Cross médias
Moments de vie



So... what?

Chiffres clés



mappy



SoLOCAL GROUP, LEADER EUROPÉEN DE LA COMMUNICATION DIGITALE LOCALE

4 800
collaborateurs

812
millions €
de chiffre d'affaires,
dont + de 80 % sur Internet

2,4 milliards
de visites
uniques par mois sur
les sites du Groupe

400 000
sites internet
conçus et gérés
par SoLocal Group

Près de
4
millions d'avis
consommateurs
disponibles sur
pagesjaunes.fr

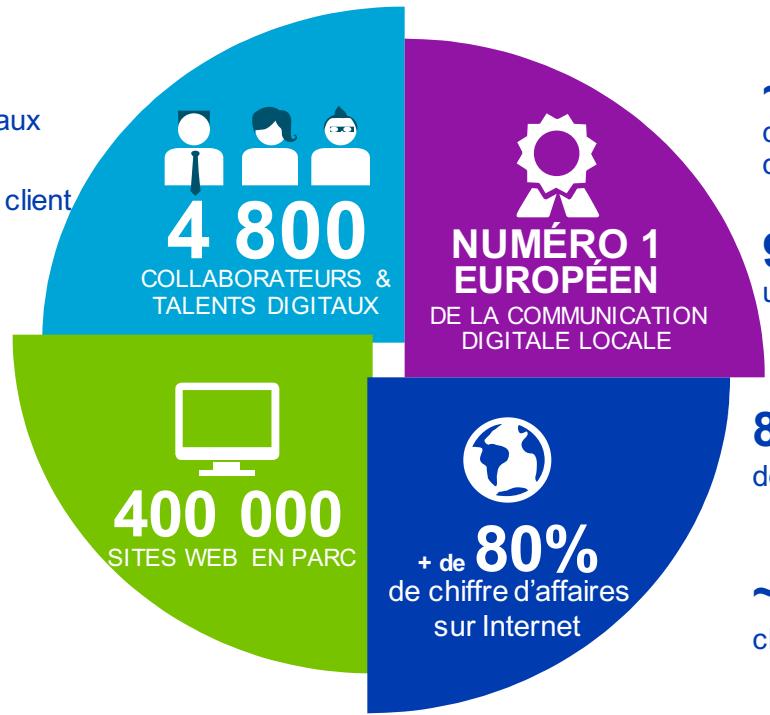
Données année 2016
Audience Médiamétrie/Netratings, année 2016

Chiffres clés

1 900 conseillers digitaux

500 conseillers service client

#1
en contenu digital
pour les PME



~30%
du marché de la publicité
digitale locale

9 Français sur 10
utilisent nos services

812 M €
de chiffre d'affaires total

~ 490 000
clients

Données année 2016

Audiences du groupe



1 770
millions de visites
dont **761 millions** sur
mobiles et tablettes



373
millions de visites
dont **171 millions** sur
mobiles et tablettes



183
millions de visites
dont **92 millions** sur
mobiles et tablettes

Autres

106
millions de visites
dont **48 millions** sur
mobiles et tablettes



Audience Médiamétrie/Netraulings, année 2016

Pages Jaunes et le Big Data – collectes de l'Audience

Pourquoi ?



Audiences de navigation : # visites, visiteurs...



Restitution détaillée aux pros



Analyse autour de l'usage : mkg, connaissance utilisateur...

Comment ?



Mediametrie

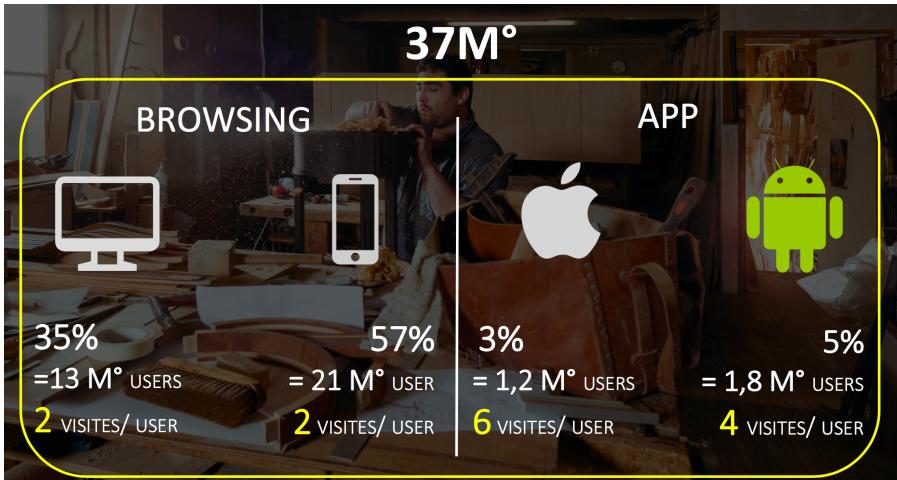
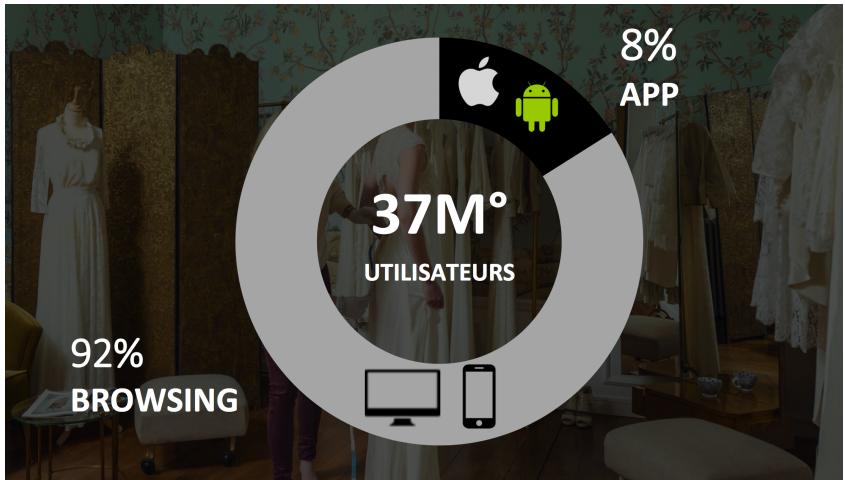


Interne
Cluster CDH
890 To de capacité stockage
Cluster Mesos
162 To de capacité stockage



Les utilisateurs sur nos médias

Répartition par support

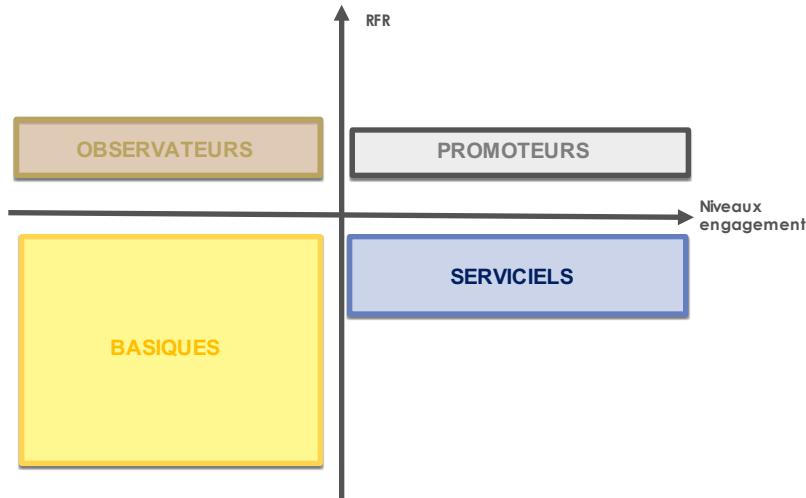


Les utilisateurs sur nos médias

Un engagement et une fréquence d'usage qui varient selon les supports

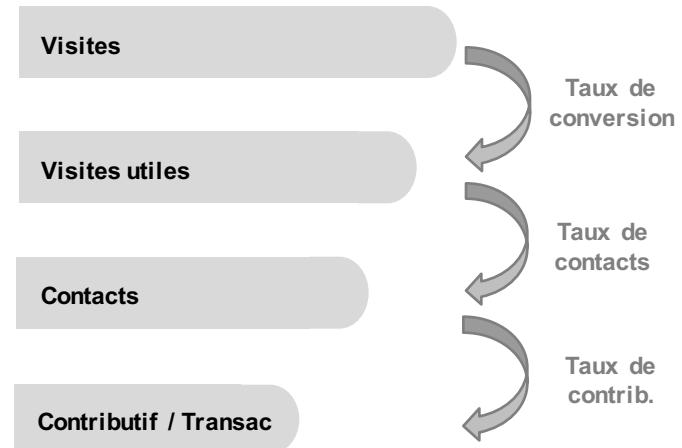
Un niveau d'engagement beaucoup plus fort sur les Applications

Le # contacts / visite varie du simple au double selon l'origine



RFR (Récence – Fréquence – Récurrence) : régularité de visite

Niveaux d'engagement : fonctionnalités utilisées (connecté, transactionnel, contributif...)



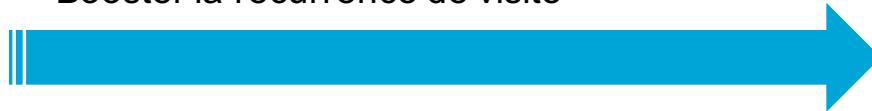
Agir pour engager et augmenter l'audience

Un objectif : transformer nos visiteurs en utilisateurs réguliers du média



SEO
Usage occasionnel
Usage coordonnées
Usage Browsing

Déplacer les usages
Activer l'usage multi device
Booster la récurrence de visite



Visite Directe
Usage régulier
Usage APP first
Compte utilisateur
Actions engageantes

Outils et cadre juridique

Data Management Platform GDPR et conséquences

Lier et normaliser la donnée

Notre volonté :

- Centraliser la donnée existantes
- Exploiter la richesse des informations
- Enrichir de nouvelles données



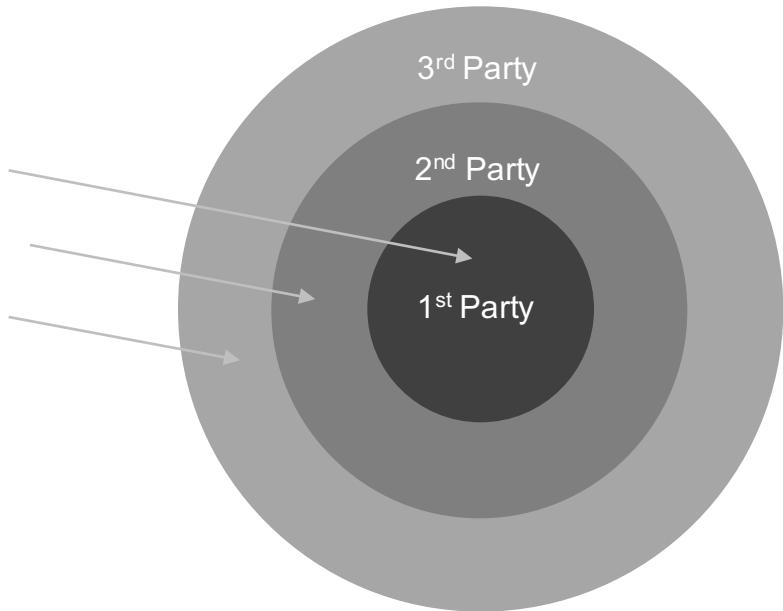
Nos enjeux :

- Améliorer la **finesse** et la **précision** de nos **segments** en enrichissant nos modèles de nouvelles informations
- Offrir de nouveaux **leviers d'actions**
- Optimiser les **temps de traitement** en centralisant la donnée et **raccourcir le TTM**

Data Management Platform

Un travail par étapes :

- « First-party data » : enrichir d'autres données interne
- « Second-party Data » : enrichir de données partenaires
- « Third-party Data » : enrichir de données extérieures



Avantages :

- Une maîtrise des coûts d'investissement
- Une démarche « test and learn » avec des cycles courts de développement
- Un gain à chaque étape

DMP – le cas PagesJaunes

	1 st Party	2 nd Party	3 rd Party	
Description	Données d'audience, collectées en interne et via des partenaires	Données du compte Favoris	Géolocalisation Ooreka Campagnes mkg	Gravity
Exemples	Parcours utilisateurs RFR Engagement	Sexe, Âge, CSP... Pros choisis en favoris	Optin Coordonnées X/Y Retours de campagne	Habitudes de connexion
Avantages	Qualité Volume Gratuité	Profil Pros favoris	Habitudes de l'utilisateur Gestion de la pression Intérêts de l'utilisateur	Centre d'intérêts plus complet
Inconvénients	Données uniquement « Jaunes » Pas d'information sur le profil	Volume plus faibles Contraintes juridiques	Dépendance partenaires Coûts (directs et de réconciliation) Contraintes juridiques	Qualité disparate Coûts (directs et réconciliation) Contraintes juridiques

Lier et normaliser la donnée

First-party data

La First-party data est stratégique. C'est la donnée dont vous êtes propriétaire (Audience, CRM, ERP...). Elle est de très bonne qualité, gratuite (hors coût des outils) et facilement accessible. Elle sert aussi de base à toutes les actions média.

Second-party data

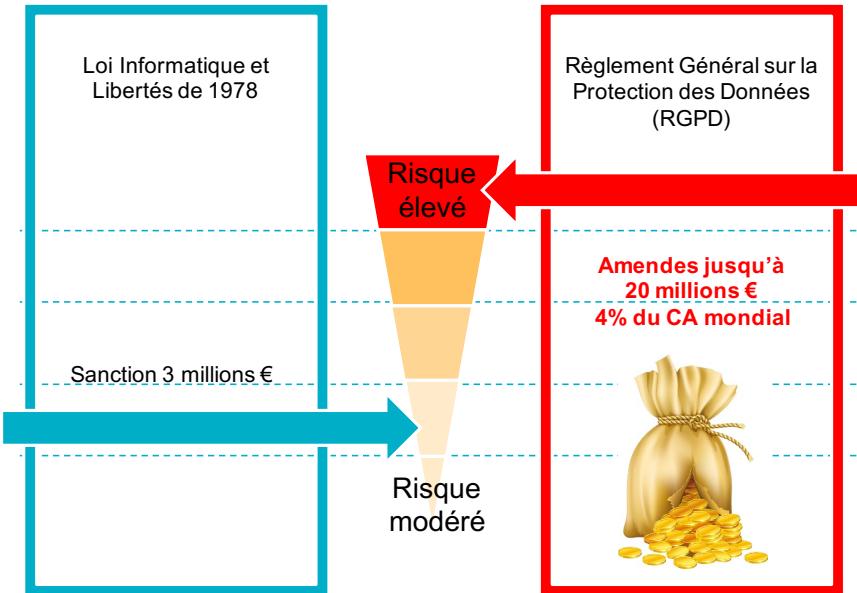
La second-party data constitue une donnée complémentaire de la donnée first-party. C'est un formidable accélérateur pour les annonceurs ayant peu de données first tout comme ceux ayant de gros volumes mais avec peu de variété.

Third-party data

La third-party data a une utilité et surtout un ROI variable, selon l'usage. Généralement couteuse, de qualité souvent disparate, elle est très liée aux cookies et peut donc s'avérer parcellaire ou peu durable. C'est aussi la donnée la plus impactée par les contraintes juridiques (GDPR).

Cadre juridique – les risques

Risque financier



Risque d'image



Cadre juridique – les questions à se poser

1/ Que vais-je faire de ces données ?
→ Déterminer les finalités du traitement

2/ De quelles données ai-je vraiment besoin ?
→ Minimiser la collecte au strict nécessaire
→ Type de données collectées ?
→ Données sensibles ?

5/ Combien de temps vais-je stocker les données ?
→ Fixer une durée de conservation

3/ Comment vais-je les collecter ?
→ Choisir des méthodes de collecte loyales et licites

4/ Qui aura accès aux données et pour en faire quoi ?
→ Restreindre les habilitations d'accès (internes-externes)



RETURN!

Agir pour accompagner

- Usages Cross médias
- Moments de vie

Le programme RETURN!

Exemple d'actions

- Segmenter les utilisateurs pour agir de façon différenciée : remarketing...
- Travailler sur du multi-verticales : connexion des activités...
- **Accompagner dans la durée : Moments de Vie**
- Agir de façon géolocalisée
- **Usage Cross Media**
- ...

1st Party

1st Party

1st Party

2nd Party

2nd Party

Moments de vie

Objectif :

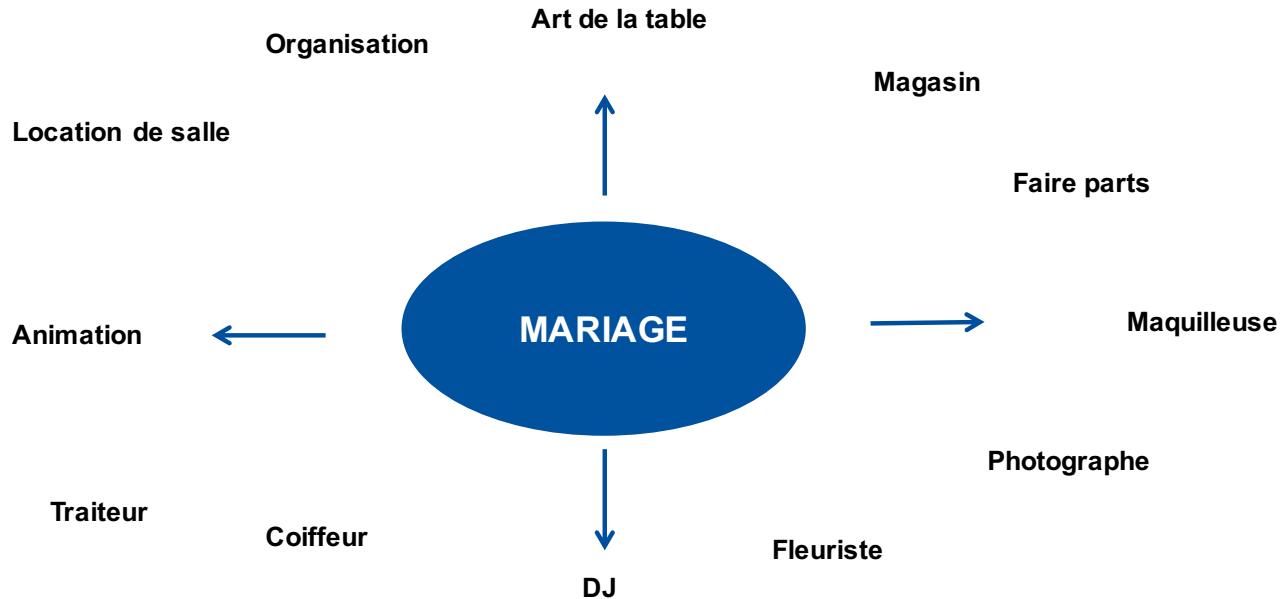
Séquencer les différentes phases d'un événement de la vie d'un utilisateur et l'accompagner dans la durée



Moments de vie - Mariage



Que tapent les utilisateurs contenant le mot mariage ?



Moments de vie - Mariage

Analyse sur les deux recherches les plus fréquentes



Organisation de mariage
Volume de visiteur : 4,4 K

2K visiteurs ont effectué une recherche sur les 2 activités



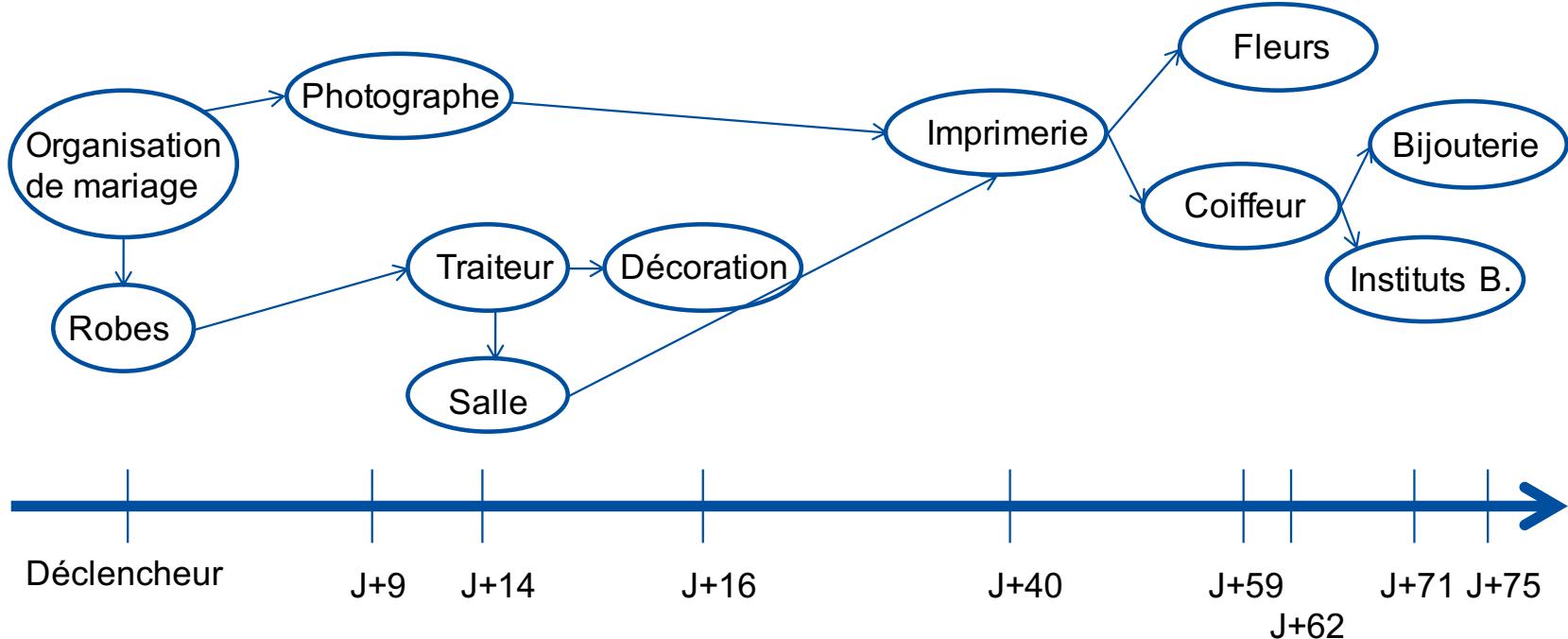
Robe de mariée
Volume de visiteur : 10,8 K

77,0% des visiteurs commencent par une recherche «Organisation de mariage », avant « Robe de mariée »
Le délai entre ces deux recherches est très court

Moments de vie - Mariage



Cycle de recherches sur la thématique



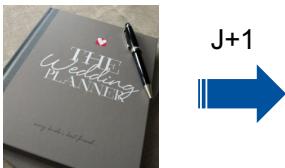
Moments de vie - Mariage



Exemple de scénario d'actions

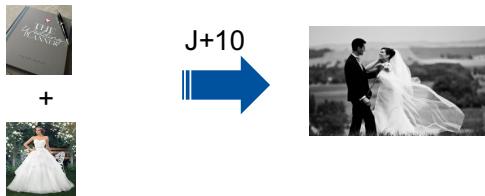
Scénario 1 : déclencheur « organisation de mariage »

Action 1 : si pas de recherche de robe de marié lors de la visite alors on pousse l'activité le plus rapidement possible



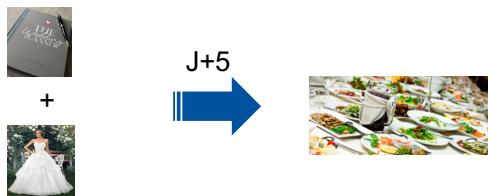
Scénario 2 : déclencheur « organisation de mariage » + « robe de mariée » dans la même visite

Action 1 : on pousse l'activité « photographe de mariage » à J+10 (si pas de recherche d'ici là)



Scénario 3 : déclencheur « organisation de mariage » + « robe de mariée » dans la même visite

Action 3 : on pousse l'activité « traiteur, organisation de réception » à J+5 (si pas de recherche d'ici là)

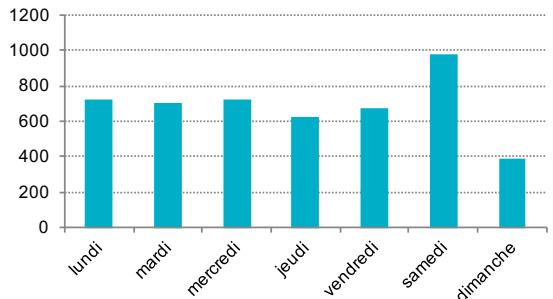


Moments de vie - Mariage

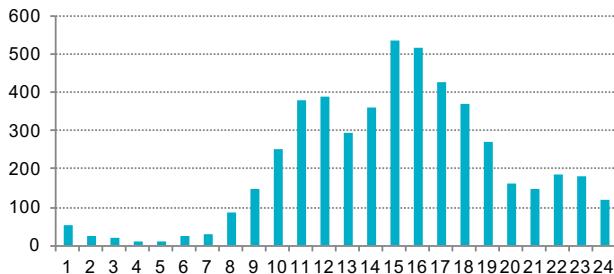


Le bon timing

Répartition dans la semaine



Répartition dans la journée



Usage cross média



Site qui propose des articles d'experts pour informer et répondre aux questions du quotidien

Par mois :

- **12,3M VU**
- **150k articles lus**



[Inscription](#) | [Connexion](#)

Maison Argent Famille Santé Droits Carrière Véhicule Entreprise

Vos questions sont entre de bonnes mains

Quel est votre sujet du moment ?

Google Recherche personnalisée



Des experts vous répondent

Poser une question



Usage cross média

Que font les utilisateurs d'Ooreka sur Pages Jaunes ?

Quels leviers activer pour booster l'audience Pages Jaunes ?



Environ 3 millions de visiteurs en commun chaque mois
sur nos 2 médias
... donc 9M de VU potentiels pour Pages Jaunes



Exemples de sous-activités :

	VU Ooreka	VU Ooreka+PJ	Potentiel
Ordinateur	57 300	10 400	46 900
Machine-à-coudre	54 300	14 600	39 700
Téléphonie	33 300	8 800	24 500

Données janvier 2017

Usage cross média – Comment faire coïncider les différents rubriques ?

1. La catégorisation Ooreka

Entretenir soi-même sa voiture

Écrit par les experts Ooreka    

Le carnet d'entretien livré par le **constructeur** vous renseigne sur les différentes opérations d'**entretien** et de **révision** à réaliser, ainsi que leur périodicité.

Vous avez plusieurs solutions pour effectuer ces opérations :

INDICATEURS GLOBAUX (AT)

X2= titre contenu ([entretenir soi-même sa voiture - ooreka](#))

X5= univers ([vehicule](#))

X8= niche ([entretien-voiture](#))

Conclusion :

S'appuyer sur la classification Ooreka donne des premiers résultats, de façon empirique.

Cependant, la généralisation à l'ensemble des articles est trop couteuse. Il faut donc trouver d'autres méthodes pour réconcilier la donnée.

Usage cross média – Comment faire coïncider les différents rubriques ?

2. Le TALn : Traitement automatique du Langage (naturel)

Entretenir soi-même sa voiture

 Écrit par les experts Ooreka    

Le carnet d'entretien livré par le **constructeur** vous renseigne sur les différentes opérations d'**entretien** et de **révision** à réaliser, ainsi que leur périodicité.

Vous avez plusieurs solutions pour effectuer ces opérations :

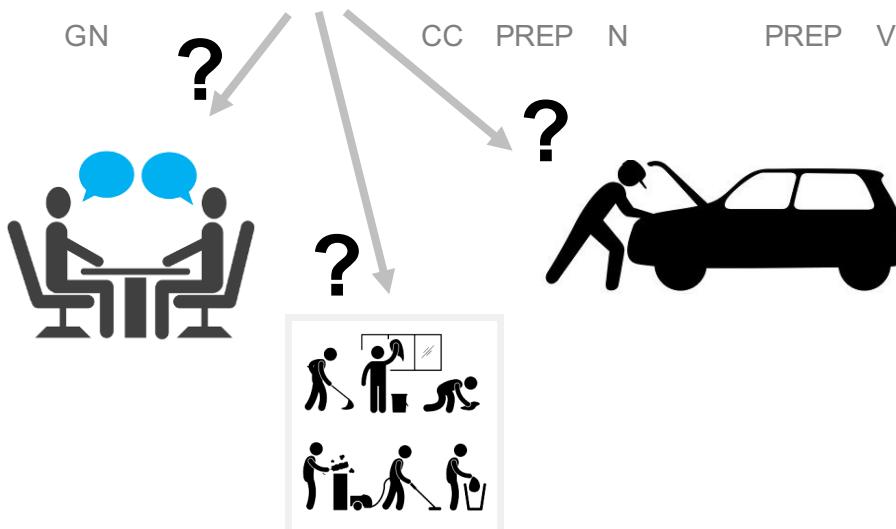
Reconnaissance d'entités nommées

- Objets textuels avec une fonction particulière
- Noms de personnes, noms d'organisation, toponymes, dates, unités monétaires, etc.
- Identification de mots clés

Usage cross média – Comment faire coïncider les différents rubriques ?

2. Le TALn : Traitement automatique du Langage (naturel)

Le carnet d'entretien livré par le constructeur vous renseigne sur les différentes opérations d'entretien et de révision à réaliser



Conclusion :

L'application d'algorithme de TALn donne des premiers résultats très encourageant sur la base de l'identification de mots clés. Cependant, certaines ambiguïtés peuvent être complexes à adresser.

Usage cross média – Comment faire coïncider les différents rubriques ?

3. Le Machine Learning (*apprentissage automatique*)

- Développement de méthodes automatiques permettant à une machine d'évoluer grâce à un processus d'apprentissage
- Effectuer des tâches qu'il est difficile ou impossible de remplir par des moyens algorithmiques plus classiques

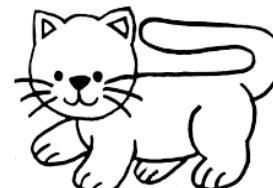
Plus simplement : répondre OUI ou NON à une question (pour peu qu'elle soit bien posée)



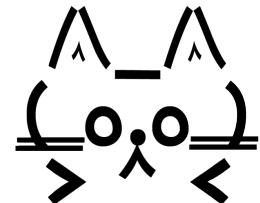
OUI



NON



OUI



?

Usage cross média – Comment faire coïncider les différents rubriques ?

Bilan

Grâce aux technologies de TALn, complétées d'algorithmes d'Apprentissage Automatique, nous pouvons réconcilier l'ensemble des articles à des sous-catégories Pages Jaunes.

Les rares cas d'ambiguïté non levés sont adressés spécifiquement sur la base de la catégorisation Ooreka.

Conclusions

Conclusions

Quelques conclusions :

- User centric c'est plus que regrouper des données utilisateur
- Centraliser et normaliser la donnée, ce n'est pas qu'une question de jointure
- Data driven et test and learn
- Et vous, vos retours ?... Et questions !

MERCI



SOLOCAL GROUP

Immeuble Citylights – Tours du Pont de Sèvres
204 Rond-point du Pont de Sèvres
92100 Boulogne Billancourt
T. 01 46 23 30 93



Guillaume GRASSET
Responsable Big Data
Connaissance User
M : 06 16 52 35 58
ggrasset@pagesjaunes.fr



PAGES JAUNES

125 bd Albert 1er – bat CCS 20714
35000
Rennes