
Christophe
Benavent

Eric
Dagiral

Caroline
Miltgen

Sarah
Medjek

RAPPORT FINAL DE RECHERCHE MESINFOS

Mars 2015





SOMMAIRE

■ 01 - INTRODUCTION

■ 02 - EXPLICATION DU PROCESSUS D'ANALYSE

■ 03 - LES DIX QUESTIONS DU RAPPORT

1. La restitution des données personnelles modifie t-elle le niveau de préoccupation à l'égard de la vie privée ?

2. La restitution des données modifie t-elle la confiance envers les entreprises détentrices des données?

3. Quelle confiance de la part des individus envers la plateforme Cozy ?

4. L'utilisation de la plateforme Cozy est-elle influencée par le niveau de préoccupation? (Privacy Concerns)

5. L'utilisation de la plateforme est-elle influencée par le niveau d'efficacité à protéger ses données personnelles? (Privacy Self Efficacy)

6. L'utilisation de la plateforme est-elle influencée par la confiance envers les entreprises? (Trust)

7. Les testeurs étaient-ils prêts à s'approprier la plateforme Cozy ?

8. L'attente de bénéfices change t-elle suite à la restitution des données personnelles ? (Benefits)

9. Les individus perçoivent-ils une valeur ajoutée quant à l'utilisation de leur espace Cozy?

10. Quelle hiérarchisation des données selon le critère de sensibilité?

■ 04 - CONCLUSION

■ 05 - ANNEXES

01

INTRODUCTION

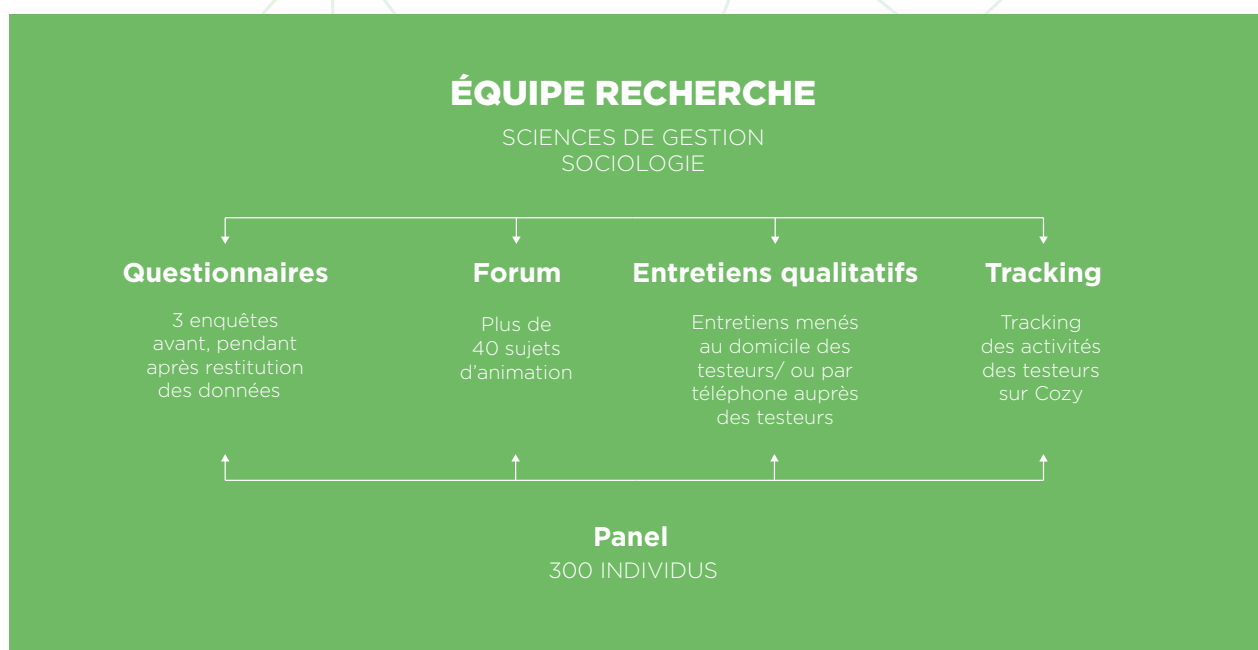
1
Retrouvez les
3 autres rap-
ports sur le site
de MesInfos :
mesinfos.fing.org

En 2013-2014, la Fing lançait une expérimentation “grandeur nature” du retour des données personnelles aux individus, engageant des partenaires divers : grandes organisations détentrices de données, acteurs technologiques, start-ups et innovateurs, chercheurs, et bien sûr des consommateurs et citoyens, qui sont les premiers concernés.

Pendant 8 mois, des grandes entreprises (banques, assurance, distribution, télécommunications) ont accepté pour la première fois de restituer à leurs clients, de manière anonyme et sécurisée, les données que ces derniers leur confient tous les jours. 300 consommateurs volontaires se sont ainsi emparés de leurs données personnelles et en ont imaginé de nouveaux usages grâce à des applications développées par des startups, étudiants et développeurs.

Une expérimentation telle que MesInfos allait certainement soulever des questions relatives à l'attitude et au comportement des testeurs vis à vis des données et des détenteurs de données. Pour y répondre, MesInfos s'est dotée d'un important dispositif de recherche, celui-ci comptant une équipe de chercheurs expérimentés. Leur objectif étant d'observer les individus composant le panel (attentes, impressions, comportement) pour récolter toutes les informations capables d'apporter des réponses aux questions soulevées au cours de l'expérimentation.

Ce rapport est le dernier d'une série de quatre rapports¹, publiés au fil de l'eau pour rendre compte des observations des chercheurs. Il intègre les principaux résultats des analyses menées durant les 8 mois de l'expérimentation.



02

EXPLICATION DU PROCESSUS D'ANALYSE

Après trois rapports préliminaires pour rendre compte des résultats des trois enquêtes menées dans le cadre de l'expérimentation Mesinfos, ce rapport vient conclure cette série, en apportant plus de détails issus notamment du tracking des activités des testeurs sur la plateforme Cozy et des entretiens menés par Eric DAGIRAL, chercheur en sociologie.

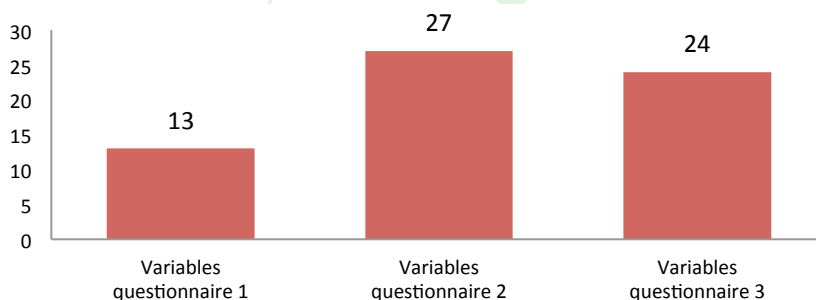
Dans ce rapport nous allons nous intéresser plus particulièrement à l'évolution de certains comportements des testeurs. Les résultats seront présentés de manière longitudinale, de manière à obtenir une confrontation des résultats en T1, T2 et T3 (correspondant aux trois questionnaires soumis aux testeurs du-

rant l'expérimentation MesInfos). Pour ce faire, la population a été réduite aux individus ayant répondu aux trois questionnaires, à savoir 56 individus.

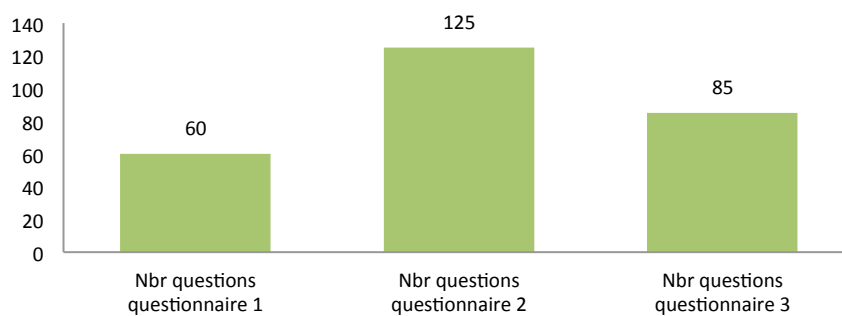
Cette population restera la même lorsqu'il s'agira de variables testées uniquement en T1, T2 ou en T2, T3, et cela dans le but de garder l'unité de notre base de calcul.

Dans le cadre des résultats relevant des traces d'usages sur la plateforme Cozy la population de base est de 180 individus, c'est à dire tous les testeurs ayant activé leurs comptes cozy et donc avec au moins une (1) trace d'usage.

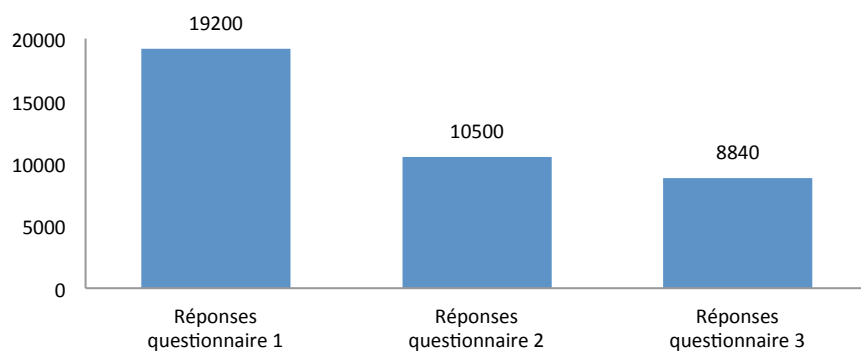
L'EXPÉRIMENTATION EN QUELQUES CHIFFRES



NOMBRE DES VARIABLES TESTÉES AU COURS
DE L'EXPÉRIMENTATION MESINFOS



**NOMBRE DE QUESTIONS POSÉES AU COURS
DE L'EXPÉRIMENTATION MESINFOS**



**NOMBRE DE RÉPONSES REÇUES AU COURS
DE L'EXPÉRIMENTATION MESINFOS**

03

LES DIX QUESTIONS DU RAPPORT

1 - LA RESTITUTION DES DONNÉES PERSONNELLES MODIFIE T-ELLE LE NIVEAU DE PRÉOCCUPATION À L'ÉGARD DE LA VIE PRIVÉE ?

Le privacy concerns fait référence à la préoccupation d'un individu concernant les éventuels dangers pouvant entourer ses données personnelles et sa vie privée. Cette préoccupation peut être liée à des facteurs externes, tels que la culture ou encore la réglementation (selon qu'elle soit stricte ou pas). Elle peut aussi être liée à des facteurs relevant de la personnalité de l'individu ou à des expériences passées (des expériences ayant un lien avec la vie privée comme le piratage d'une boîte mail, en anglais *privacy experience*).

Il existe plusieurs échelles pouvant mesurer le niveau de préoccupation à l'égard de la vie privée, et dans le cadre de l'expérimentation MesInfos c'est l'échelle IUIPC (*Internet User Information Privacy Concerns*) qui a été utilisée du fait de son adéquation à un contexte numérique.

La première enquête de l'expérimentation (lancée en octobre 2013) nous a révélé que notre panel, composé alors de 320 individus, avait un niveau de préoccupation élevé. Nous avons alors émis l'hypothèse que ce ni-

veau de préoccupation allait changer/évoluer suite à la restitution des données.

Lors de la seconde enquête cette variable a été testée de nouveau et une comparaison T1 Vs T2 a été menée.

Cette comparaison des moyennes n'a démontré qu'une légère évolution non significative. L'explication de ce « léger » changement pouvait résider dans le fait que nous étions face à un panel avec un niveau élevé de conscience (conscient de la collecte et de l'utilisation de leurs données) et donc suite à la restitution des données, il n'y a pas eu de choc susceptible de provoquer un changement de comportement ou d'attitude.

Pour ce derniers rapport nous avons effectué une comparaison des moyennes en T1, T2, et T3, là encore la différence n'a été que très peu significative. Pour aller plus loin dans l'analyse, nous avons aussi effectué des tests de Student.

Pour information : Un test de Student est un test statistique utilisé pour mesurer l'hypothèse d'égalité de deux variables.

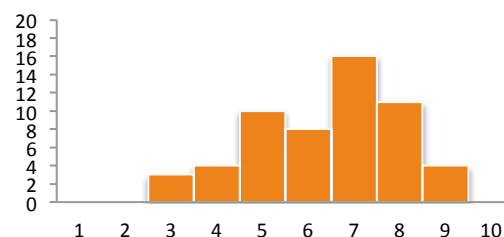
Ce test part de l'hypothèse que deux variables (ou deux moyennes d'une même variable) sont égales. Si le T de Student est supérieur à 1.96 ou inférieur à -1.96 nous rejetons l'hypothèse d'égalité (il y a différence entre les deux moyennes ou les deux variables), si le T de Student se situe entre 1.96 et -1.96 nous acceptons l'hypothèse d'égalité (il n'y a pas de différence).

Comme le démontrent les résultats ci-dessous, le T de Student est de 0,84 **entre le T1 et le T2** (soit une valeur se situant entre 1.96 et -1.96) ce qui signifie que les deux moyennes ne sont pas significativement différentes (pas de modification notable du niveau de préoccupation). En revanche, le T de Student est de -2.95 **entre le T2 et le T3**, cela indique que l'augmentation de la moyenne du niveau de préoccupation est assez significative : en somme, l'inquiétude des testeurs concernant leur vie privée a augmenté légèrement entre le T2 et le T3.

La différence des moyennes entre le T1 et le T3 est importante mais le T de Student indique -1.85 (soit une valeur entre 1.96 et -1.96) ce qui signifie que cette différence n'est pas assez importante au regard du teste de Student (cela peut être expliqué par un faible écart type).

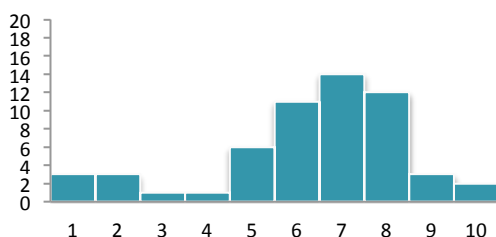
Ces résultats nous amène à dire que le niveau de préoccupation est resté constant tout long de l'expérimentation (une moyenne de 6/10) mais avec des fluctuations. Cette préoccupation a atteint son niveau le bas en T2 et son niveau le plus élevé en T3 (ce qui explique la différence significative entre ces deux périodes), nous émettons comme hypothèse que les participants ont pris conscience de la récolte massive de données, introduisant ainsi une légère augmentation du niveau de préoccupation.

MOYENNE : 6.53 ET 1.62



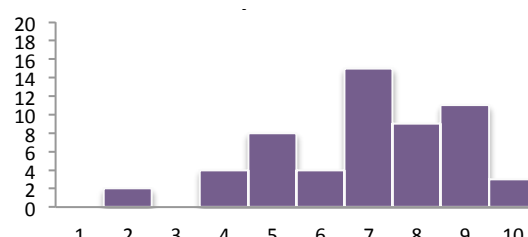
PRIVACY CONCERNS T1

MOYENNE : 6.321 ET 2.107



PRIVACY CONCERNS T2

MOYENNE : 6.948 ET 1.838



PRIVACY CONCERNS T3

CALCUL DU T DE STUDENT : (MOYENNES PC T1, T2, T3)

Data : Préoccupation T1, Préoccupation T2 $t = 0.84$, $p\text{-value} = 0.401$
Data : Préoccupation T2, Préoccupation T3 $t = -2.95$, $p\text{-value} = 0.004$
Data : Préoccupation T1, Préoccupation T3 $t = -1.85$, $p\text{-value} = 0.069$

2 - LA RESTITUTION DES DONNÉES PERSONNELLES MODIFIE T-ELLE LA CONFIANCE ENVERS LES ENTREPRISES DÉTENTRICES DES DONNÉES ?

La confiance est sans doute une des variables sur lesquelles les chercheurs ont porté une attention toute particulière. Dans le contexte de l'expérimentation MesInfos, la confiance peut être définie comme la conviction d'un individu que l'entreprise dont il est client protège ses données personnelles et les utilise à bon escient.

La confiance a été mesurée lors de la première et de la seconde enquête. Comme indiqué sur les schémas en face, le niveau de confiance du panel lors de la première enquête était moyen (avec une moyenne de 5.625/10). Lors de la seconde enquête le niveau de confiance est resté assez similaire mais avec une légère augmentation de l'axe "confiance basse", mais cette augmentation est-elle significative? Pour cela nous nous sommes appuyés sur un test de Student.

Les résultats du test de Student indiquent un T à 1.68 soit une valeur se situant entre 1.96 et -1.96, ce qui signifie que le changement de moyenne survenu lors de l'enquête N°2 n'est pas assez significatif pour qu'on puisse le prendre en compte. **Ainsi le niveau de confiance n'a pas été impacté par la restitution des données.**

En revanche, nous pouvons noter une nette augmentation de la confiance envers la plateforme Cozy (voir Q3).

En effet, dès la seconde enquête, les chercheurs ont voulu chercher à évaluer la confiance des individus envers la plateforme. En définitive, c'est bien sur cet espace que toutes les données des testeurs étaient stockées et éventuellement utilisées (via des applications par exemple) et de ce fait la confiance envers Cozy paraissait être toute aussi intéressante et importante à mesurer que celle envers les entreprises détentrices des données. Ainsi lors de l'enquête N°2, le niveau de confiance général envers la plateforme Cozy était moyen, d'ailleurs très similaire à la confiance envers les entreprises, avec une

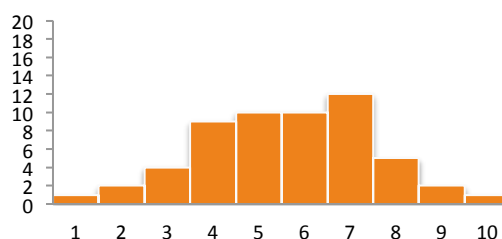
moyenne de 5.882. En revanche, lors de la troisième enquête ce niveau de confiance a largement augmenté avec une moyenne de 7.862.

Cette augmentation a été confirmée par le test de Student qui nous indique un T de -7.517, une valeur bien au dessous de -1.96, les deux moyennes T2 Vs T3 ne sont pas égales, il y a donc une différence significative.

Ce haut niveau de confiance pourrait être directement lié à la perception des individus concernant la plateforme Cozy, à savoir une plateforme sécurisée visant à leur redonner le contrôle sur leurs données personnelles et pouvant être quittée à tout moment (avec une certitude que les données seront supprimées à leur demande)

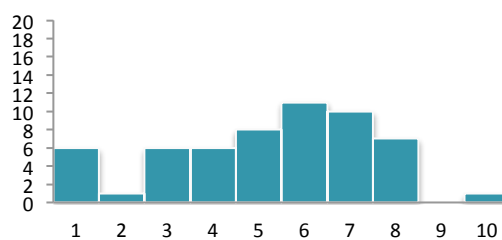
CALCUL DE MOYENNES

MOYENNE : 5.66, ET : 1.88



CONFIANCE T1

MOYENNE : 5.228 ET 2.210



CONFIANCE T2

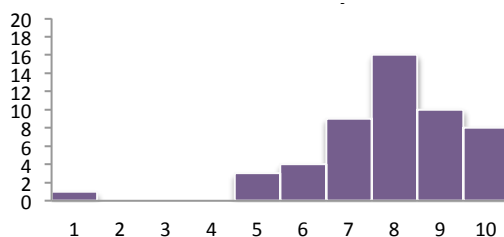
T DE STUDENT (MOYENNE CONFIANCE T1, T2)

Data : Confiance T1, Confiance T2
t = 1.68, p-value = 0.097

3 - LE NIVEAU DE CONFIANCE DANS LA PLATEFORME COZY

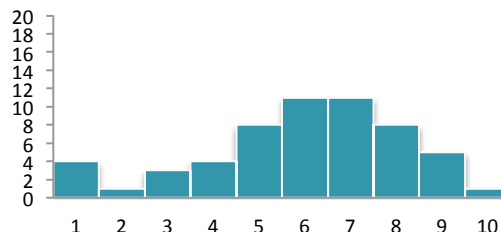
CALCUL DE MOYENNES

MOYENNE : 7.86, ET : 1.68



CONFIANCE T1

MOYENNE : 5.88, ET : 2.23



CONFIANCE T2

T DE STUDENT (MOYENNE CONFIANCE T1, T2)

Data : Confiance dans le plateforme T2, Confiance dans la plateforme T3)
 $t = -7.5175$, $df = 50$, $p\text{-value} = 9.334e-10$

4 - L'UTILISATION DE LA PLATEFORME COZY EST-ELLE INFLUENCÉE PAR LE NIVEAU DE PRÉOCCUPATION À L'ÉGARD DE LA VIE PRIVÉE?

Tout l'enjeu de l'expérimentation MesInfos était de restituer aux testeurs leurs données personnelles et leur permettre de les réutiliser pour en faire ce qui a du sens pour eux. La restitution des données s'est faite via la plateforme Cozy dans le but d'assurer la sécurité des données et de fournir les fonctionnalités et les applications nécessaires à leur réutilisation.

Le tracking des activités des testeurs sur la plateforme Cozy nous a démontré que 180 individus (sur les 320 initiaux) ont activé leur

espace Cozy. La moyenne de traces d'usage est de 14 traces (le min = 1 trace, la max= 267 traces)

L'usage de la plateforme a été moins important que ce qui était attendu, de plus l'utilisation a été sous forme de "one shot", c'est à dire que la majorité des testeurs ne s'est connectée qu'une seule fois et a testé différentes applications. Mais est-ce que l'utilisation de la plateforme a été influencée par d'autres variables comme la confiance ou la préoccupation ?

L'Utilisation de la plateforme était-elle influencée par le niveau de préoccupation à l'égard de la vie privée ?

Afin de répondre à cette question nous avons réparti les testeurs en deux groupes ;

- Groupe 1 : individus ayant un niveau de préoccupation bas et moyen.
- Groupe 2 : individus ayant un niveau de préoccupation élevé.

Nous avons procédé par la suite à la confrontation du niveau de préoccupation et de l'utilisation de la plateforme selon trois groupes d'utilisateurs :

- Groupe A : individus ayant de 1 à 20 traces d'usage.
- Groupe B : individus ayant de 21 à 40 traces d'usage.
- Groupe C : individus ayant plus de 41 traces d'usage.

Comme le démontre le schéma ci-dessous, cette répartition nous permet de connaître l'utilisation de la plateforme selon le niveau de préoccupation mais ne nous dit pas s'il y a corrélation. Pour cela nous avons procédé à un test de Chi2.

Pour information : Un test de Chi2 ou Khi2 est un test statistique qui permet de mesurer l'indépendance de deux variables différentes et dont l'observation est synthétisée sous la forme d'un tableau de contingence. Le test de Chi2 part de l'hypothèse selon laquelle les deux variables sont indépendantes. Si la "P value" est supérieure à 0,05 nous acceptons l'indépendance des variables (elles ne sont pas liées) si la "P value" est inférieure à 0,05 nous rejetons l'indépendance des variables (elles sont liées).

Tableau des valeurs observées	PC (bas moyen)	PC élevé	
01>>20	105	46	151
21>>40	9	4	13
41>>	10	6	16
	124	56	180

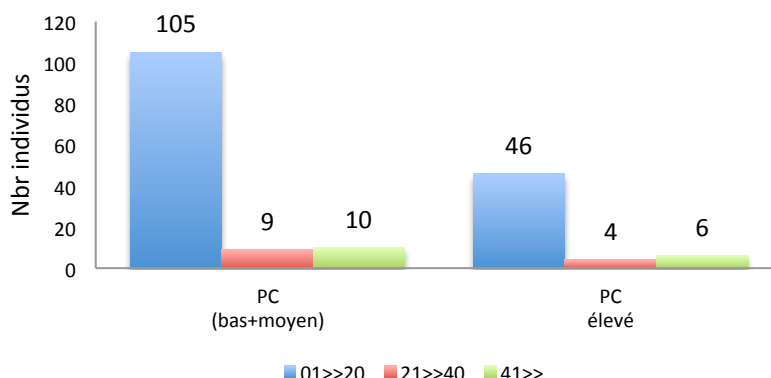
Tableau des valeurs attendues	PC (bas+moyen)	PC élevé
	0,6889	0,3111
01>>20	104,0222	46,97778
21>>40	8,9556	4,04444
41>>	11,0222	4,97778

P-value	0,854
---------	-------

Comme nous pouvons le voir dans les tableaux ci-dessus la P value du chi2 (usage Vs Préoccupation) est de 0.854 soit une valeur supérieure à 0.05, l'indépendance des variables est acceptée, **l'utilisation de la plateforme n'est pas influencée par le Privacy concerns.**

CALCUL DU CHI2

UTILISATION DE LA PLATEFORME SELON NIVEAU DE PRÉOCCUPATION



5 - L'UTILISATION DE LA PLATEFORME EST-ELLE INFLUENCÉE PAR LE NIVEAU D'EFFICACITÉ À PROTÉGER SES DONNÉES PERSONNELLES? (PRIVACY SELF EFFICACY)

Pour répondre à cette question nous avons de nouveau réparti les testeurs en deux groupes ;

- Groupe 1 : individus ayant un niveau d'efficacité bas et moyen.
- Groupe 2 : individus ayant un niveau d'efficacité élevé.

Nous avons procédé par la suite à la confrontation du niveau d'efficacité et de l'utilisation de la plateforme selon trois groupes d'utilisateurs :

- Groupe A : individus ayant de 1 à 20 traces d'usage.
- Groupe B : individus ayant de 21 à 40 traces d'usage.
- Groupe C : individus ayant plus de 41 traces d'usage.

Comme le démontre le schéma ci-dessous, cette répartition nous permet de connaître l'utilisation de la plateforme selon le niveau d'efficacité mais ne nous dit pas s'il y a corrélation, Pour cela nous avons procédé à un test de Chi2

Tableau des valeurs observées

	PSE (bas moyen)	PSE élevé	
01>>20	111	40	151
21>>40	12	1	13
41>>	8	8	16
	131	49	180

Tableau des valeurs attendues

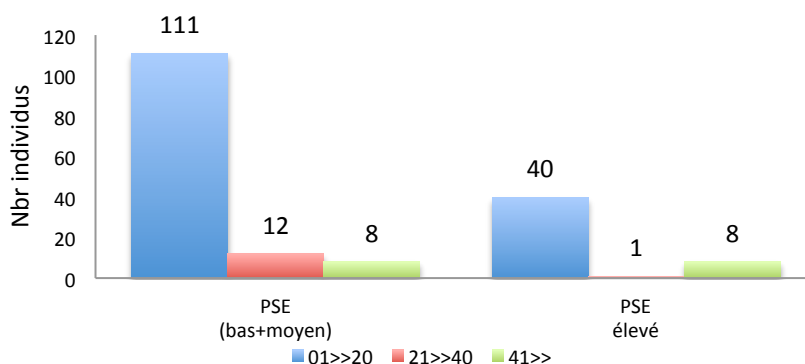
	PSE (bas+moyen)	PSE élevé
	0,7278	0,2722
01>>20	109,8944	41,10556
21>>40	9,4611	3,53889
41>>	11,6444	4,35556

P-value	0,0345
---------	--------

Comme nous pouvons le voir dans les tableaux ci-dessus la P value du chi2 (usage Vs efficacité) est de 0.0345 soit une valeur inférieure à 0.05 l'indépendance des variables est rejetée, **l'utilisation de la plateforme est liée au sentiment d'efficacité à protéger ses données personnelles.**

CALCUL DU CHI2

UTILISATION DE LA PLATEFORME SELON NIVEAU DE PRÉOCCUPATION



6 - L'UTILISATION DE LA PLATEFORME EST-ELLE INFLUENCÉE PAR LA CONFIANCE ENVERS LES ENTREPRISES? (TRUST)

Pour répondre à cette question nous avons réparti les testeurs en trois groupes

- Groupe 1 : individus ayant un niveau de confiance.
- Groupe 2 : individus ayant un niveau de confiance moyen.
- Groupe 3 : individus ayant un niveau de confiance élevé.

Nous avons procédé par la suite à la confrontation du niveau de préoccupation et de l'utilisation de la plateforme selon trois groupes d'utilisateurs :

- Groupe A : individus ayant de 1 à 20 traces d'usage.
- Groupe B : individus ayant de 21 à 40 traces d'usage.
- Groupe C : individus ayant plus de 41 traces d'usage.

Comme le démontre le schéma ci-dessous, cette répartition nous permet de connaître l'utilisation de la plateforme selon le niveau de confiance mais ne nous dit pas s'il y a corrélation, pour cela nous avons procédé à un test de Chi2.

Tableau des valeurs observées

	TR (bas)	TR (moyen)	TR (élevé)	
01>>20	23	113	15	151
21>>40		13		13
41>>	2	12	2	16
	25	138	17	180

Tableau des valeurs attendues

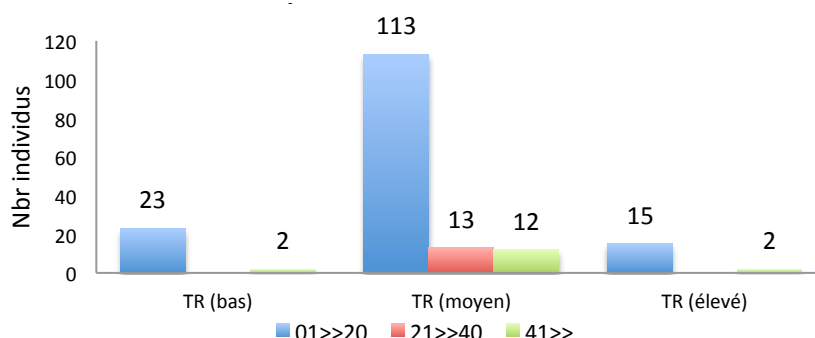
	TR (bas)	TR (moyen)	TR (élevé)
	0,1389	0,7667	0,0944
01>>20	20,972	115,766	14,261
21>>40	1,805	9,966	1,227
41>>	2,222	12,266	1,511

P-value	0.3493
---------	--------

Comme nous pouvons le voir dans les tableaux ci-dessus la P value du chi2 (usage Vs confiance) est de 0.349 soit une valeur supérieure à 0.05, l'indépendance des variables est acceptée, l'utilisation de la plateforme n'est pas influencée par la confiance envers les entreprises.

CALCUL DU CHI2

UTILISATION DE LA PLATEFORME SELON LE NIVEAU DE CONFIANCE

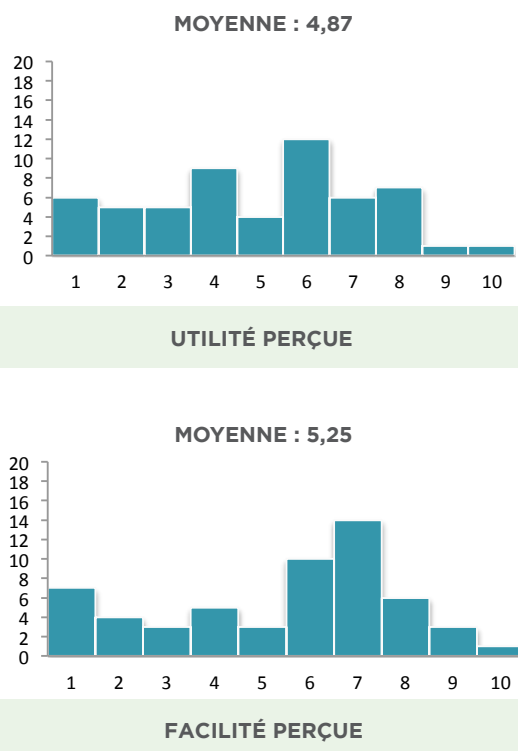


7 - LES TESTEURS ÉTAIENT-ILS PRÊTS À S'APPROPRIER LA PLATEFORME COZY ?

Comme nous l'avons mentionné plus haut, l'utilisation de la plateforme est restée assez modeste et était influencée par le niveau de privacy self efficacy (efficacité à protéger ses données personnelles et sa vie privée). La question ici est de savoir si les testeurs étaient prêts à utiliser la plateforme. En effet, avant même de parler de l'utilisation de la plateforme, n'est-il pas intéressant de se pencher sur la propension des testeurs à utiliser cette plateforme ? Étaient-ils disposés à utiliser et s'approprier la plateforme ? Pour répondre à cette question nous allons faire appel à l'échelle appelée TAM *Technology Acceptance Model*.

Cette échelle est la référence en matière d'adoption et d'acceptation d'une nouvelle technologie, elle vise à mesurer la propension (*readiness*) des individus à adopter une technologie. Elle se base sur leur perception de l'utilité et de la facilité d'utilisation de la technologie en question (l'utilité ayant un impact plus fort).

Lors de la seconde enquête cette échelle a été incluse dans le questionnaire, ainsi l'utilité perçue et la facilité d'utilisation perçue (de la plateforme Cozy) ont été mesurées. Voici les résultats :



Les résultats ci-dessus nous démontrent que les testeurs ont une perception de l'utilité et de la facilité d'utilisation de la plateforme qui est faible.

Avec des moyennes de 4.87/10 pour l'utilité et 5.25/10 pour la facilité d'utilisation, ces résultats nous indiquent, selon l'échelle TAM, **une faible propension à utiliser la plateforme Cozy.**

8 - L'ATTENTE DE BÉNÉFICES CHANGE-T-ELLE SUITE À LA RESTITUTION DES DONNÉES PERSONNELLES ? (BENEFITS)

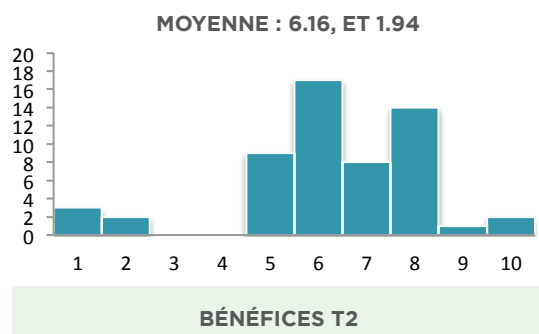
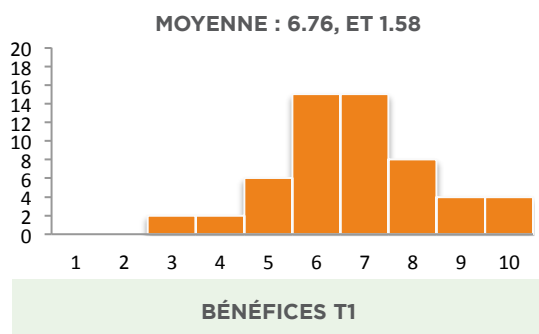
Motiver les individus à partager leurs données personnelles contre l'obtention d'un avantage est une pratique très répandue. En effet, avant de partager ses données, l'individu évalue les coûts et les avantages, si ces derniers sont

jugés intéressants, l'individu acceptera plus facilement de divulguer ses données. Les avantages (ou bénéfices attendus) peuvent être extrinsèques à l'individu (comme le gain de temps) ou intrinsèques (comme le divertissement).

La première enquête de MesInfos nous a démontré que le niveau d'attente de bénéfices au sein du panel était assez élevé avec une moyenne de 6,76/10, Mais est ce que ce niveau d'attente a été impacté par la restitution des données ?

Lors de la seconde enquête la moyenne de cette variable était de 6,16, enregistrant ainsi

une diminution, mais cette dernière est-elle assez significative pour être prise en compte? Pour cela nous avons procédé à un test de Student.



T DE STUDENT
(MOYENNES BENEFITS T1, BENEFITS T2)

Data = Bénéfices T1, Bénéfices T2
 $t = 2.05$, $p\text{-value} = 0.044$

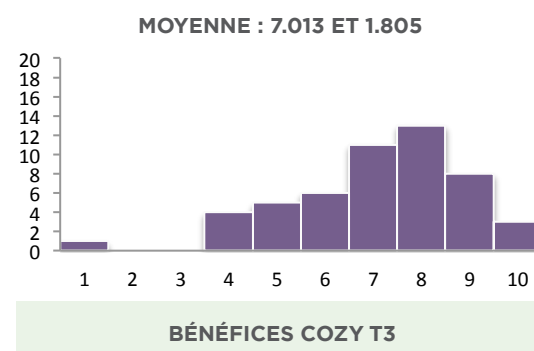
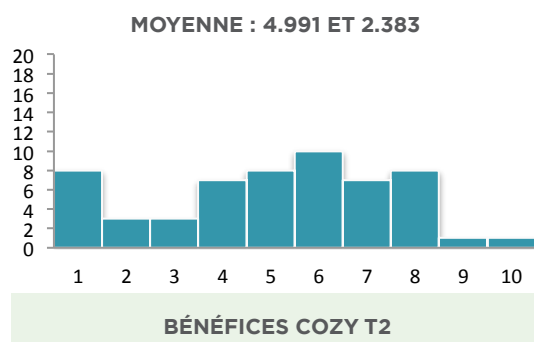
Entre T1 et T2, le T de Student est de 2.05 (donc supérieur à 1.96) l'hypothèse d'égalité est rejetée, la diminution enregistrée est significative. De même, l'attente de bénéfices de la part de la plateforme Cozy a enregistré une augmentation significative en T3. En effet, comme le démontrent les Schémas ci-après, la moyenne est passée de 4.99/10 en T2 à 7.01/10 en T3. De plus, le test de Student indique un T de -7.89 soit une valeur inférieure à -1,96, l'hypothèse d'égalité est rejetée, la différence observée est significative et peut être prise en compte.

Dans les deux cas l'attente de bénéfices change suite à la restitution des données. L'augmentation de cette attente est plus importante envers Cozy, cela peut être expliqué

par la promesse même de la plateforme, qui permet aux individus de tirer des bénéfices de leurs propres données personnelles.

COMPARAISON AVEC L'ATTENTE DE BÉNÉFICES DE LA PART DE LA PLATEFORME COZY.

CALCUL DES MOYENNES



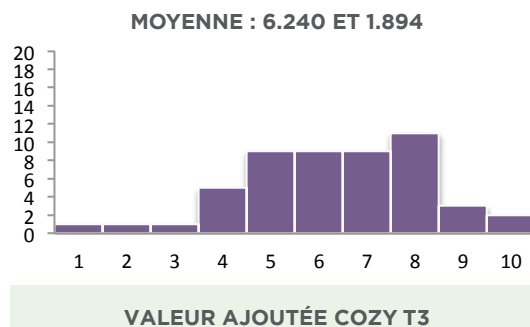
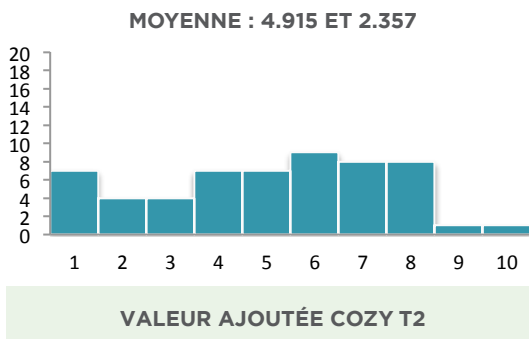
T DE STUDENT (MOYENNES BENEFITS T1, BENEFITS T2)

Data : Bénéfices T2, Bénéfices T3
 $t = -7.8964$, $p\text{-value} = 2.406e-10$

9 - LES INDIVIDUS PERÇOIVENT-ILS UNE VALEUR AJOUTÉE QUANT À L'UTILISATION DE LEUR ESPACE COZY?

La plateforme Cozy devait permettre aux testeurs de réutiliser leurs données personnelles et en tirer profit à leur tour. Cette plateforme devait aussi apporter des services de qualité, innovants et utiles avec une réelle valeur ajoutée. Nous nous intéressons ici à comprendre si les testeurs, utilisateurs de la plateforme, ont eux perçu cette valeur ajoutée.

Lors de la seconde enquête, la perception de la valeur ajoutée était assez faible avec une moyenne de 4.15, cette moyenne est passée à 6.24 lors de la troisième enquête. Pour savoir si cette évolution est significative, un test de Student a été effectué. Celui-ci indique un T de -5.88 soit une valeur inférieure à -1.96 et donc l'hypothèse d'égalité est rejetée, **l'augmentation de moyenne que nous avons observée est donc significative et peut être prise en compte.** Cette augmentation de la perception de la valeur ajoutée peut être directement liée à l'utilisation de la plateforme. En effet, lors de la seconde enquête les testeurs n'avaient accès à la plateforme que depuis peu de temps. Lors de la troisième enquête, les testeurs s'étaient plus familiarisés avec la plateforme et ses fonctionnalités, et avaient ainsi pris conscience de toute la valeur ajoutée qu'elle pouvait leur apporter.



T DE STUDENT

Data : Valeur ajoutée T2, Valeur ajoutée T3

$t = -5.883$, $p\text{-value} = 3.325e-07$

10 - QUELLE HIÉRARCHISATION DES DONNÉES SELON LE CRITÈRE DE SENSIBILITÉ?

Eric Dagiral, chercheur en sociologie, s'est rendu au domicile de testeurs pour avoir leurs retours sur l'expérimentation MesInfos. À cette occasion il a demandé aux testeurs de hiérarchiser les données des moins sensibles aux plus sensibles. Selon les testeurs, les données pourraient être hiérarchisées comme ceci :

- **Les données « pas très sensibles » :** Nom, Prénom, Age, Adresse
- **Les données « relativement sensibles » :** Traces de navigation, Carte de fidélité (mais là il s'agit plus de données de foyer, plus NosInfos que MesInfos)
- **Les données « très sensibles » :** Géolocalisation, Carte bleue, Santé (notamment les données QS)

Néanmoins, il est important de préciser que pour de nombreux testeurs ce ne sont pas les données en elles-mêmes qui doivent être qualifiées de sensibles, mais leur place dans l'expérience des individus corrélativement aux situations de vie, en lien avec un processus de découverte et d'apprentissage (parfois d'expérimentation) de ce que sont les données numériques. Cette expérience personnelle, est à la fois profondément encadrée dans des épreuves individuelles et familiales d'une part, et dans des controverses sociétales d'autre part (dans l'enquête MesInfos le «moment» Snowden/NSA).

Enfin, il semble donc ressortir des entretiens, que pour les plus familiers de ces controverses sociétales, émerge une conscience solide du fait que ce ne sont pas des types de données qui par elles-mêmes (prises isolément) sont plus ou moins sensibles, mais que ce sont les pratiques d'analyse et de croisements de données hétérogènes qui sont génératrices de degrés de sensibilité. En effet, les données de géolocalisation sont revenues dans plusieurs entretiens dans ce sens, pointant que leur récolte et leur utilisation pouvaient s'avérer aussi inoffensives, traitées par certaines sociétés, que liberticides dans d'autres contextes.

04

CONCLUSION

L'expérimentation MesInfos a été riche en enseignements tant du point de vue méthodologique que du point de vue des résultats obtenus.

Tout d'abord, elle nous a démontré qu'il n'y a pas eu de choc au sein du panel face à la restitution des données personnelles. Les testeurs étant tout à fait conscients de la collecte et de l'utilisation des données par les organisations, cette restitution n'a pas mené à des changements radicaux de comportements ou d'attitudes. Cependant, nous avons tout de même observé de légères évolutions de certaines variables. En effet, comme nous avons pu le constater, mettre à disposition des individus leurs propres données personnelles a eu un impact (même s'il reste léger) sur leur niveau de préoccupation à l'égard de la vie privée et sur leur niveau d'attente de bénéfices de la part des entreprises détentrices de données.

MesInfos nous apprend également que le partage des données et l'utilisation d'une plateforme de gestion de données personnelles (dans notre cas la plateforme Cozy) est fortement liées à la confiance en soi (l'efficacité à protéger ses données personnelles et sa vie privée). En effet, lors de notre première enquête nous avons déjà mis en lumière la corrélation existante entre l'efficacité et le partage des données personnelles. Dans ce rapport final, nous avons une nouvelle fois démontré, grâce aux tests de χ^2 , le lien existant entre l'efficacité et l'utilisation de la plateforme Cozy. De ce fait, pour les entreprises qui souhaitent emprunter la voie de la restitution des données, cela suppose qu'elles

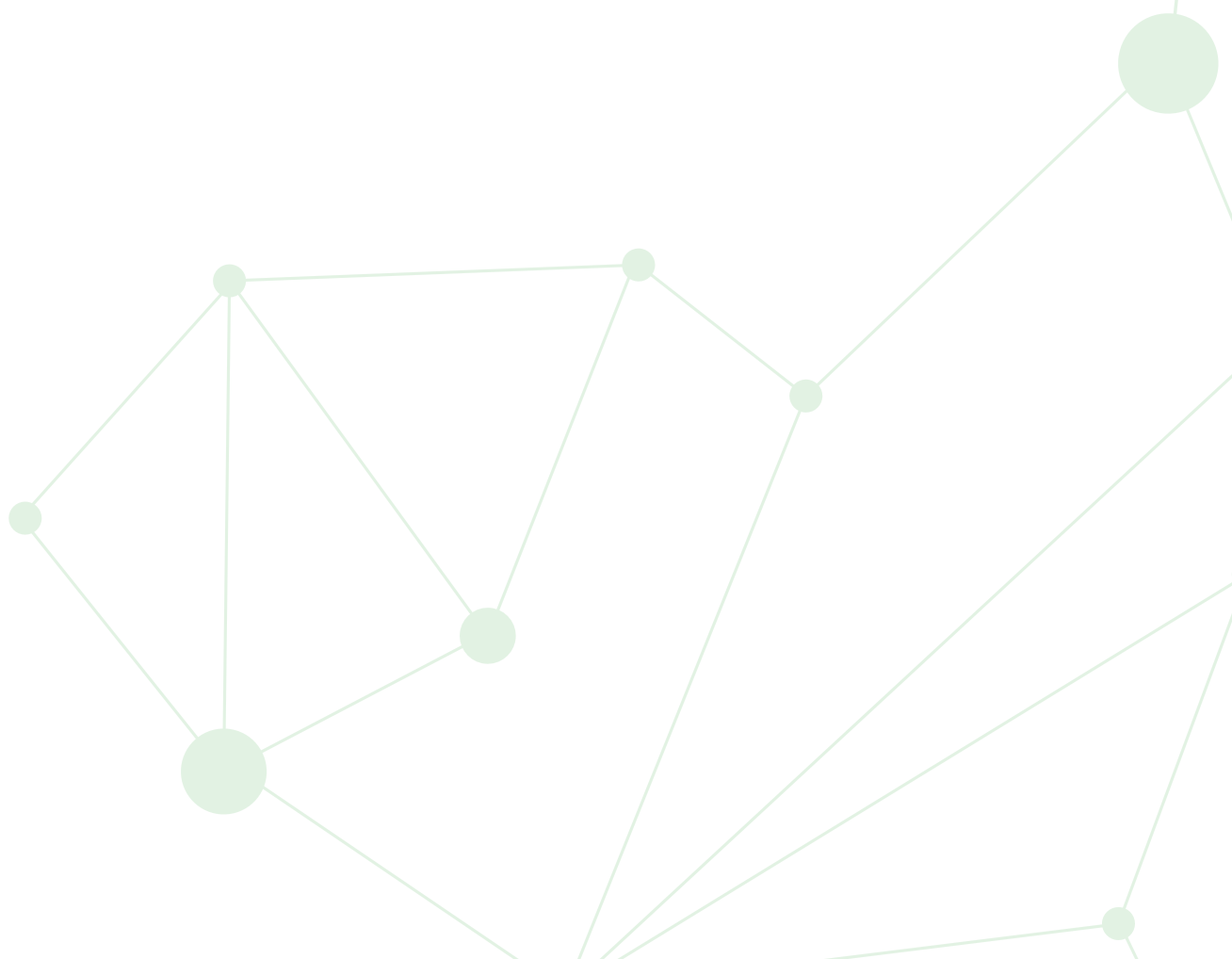
doivent prendre en compte non seulement la confiance qu'elles suscitent chez leurs clients/usagers, mais aussi la confiance de ces derniers en eux-mêmes et en leurs capacités à protéger leur vie privée.

Enfin, l'expérimentation s'avère tout à fait positive pour les partenaires du projet. En effet, 67% des participants à la troisième enquête émettent un avis positif sur l'expérimentation et considèrent mieux se connaître, et 70% d'entre eux indiquent se sentir plus en capacité de protéger leurs données personnelles. Dans le même temps, alors que la confiance globale envers les entreprises détentrices de données stagne à un niveau moyen, les partenaires de MesInfos bénéficient d'un fort intérêt et d'un attachement significatif de la part des testeurs. Ces derniers ont répondu à près de 80% se sentir plus engagés envers les partenaires du projet qui ont su faire preuve de transparence.

Cette expérimentation tout à fait singulière s'est achevée en mai 2014 mais elle a ouvert la voie à de nouveaux défis économiques, juridiques et techniques. Des pistes qui vont permettre d'identifier les modèles économiques, de mettre en place un cadre juridique et de créer un environnement technique sécurisé et approprié de ce que nous appelons désormais le Self-Data.

05

ANNEXES



Expérimentation MesInfos

FING

Volet d'enquête qualitative

Synthèse des principaux résultats (version intermédiaire), décembre 2014

Par Éric Dagiral (Université Paris Descartes, CERLIS), avec la contribution de l'équipe recherche de MesInfos et la participation de Vinciane Zabban

1. Présentation du cadre méthodologique, et de quelques incidences

Le volet qualitatif de l'enquête s'est déroulé en deux temps. Un premier moment en amont même du dispositif d'expérimentation MesInfos avait été mené par la sociologue Vinciane Zabban (Université Paris 13, EXPERICE) autour de l'été 2012, à travers la passation de 6 entretiens (enquête-e-s adultes résidant en région parisienne) visant à explorer la faisabilité d'une approche par entretiens autour d'une question à la fois si vaste, et si technique dans sa problématisation, que celle des « données personnelles ». Ces premiers entretiens soulignaient combien une telle expression ne faisait pas nécessairement sens pour les individus, et qu'il devenait important de privilégier des entrées par des pratiques - et des façons de les nommer - plus répandues : administration et consommation ; agenda et déplacement ; sport et santé ; loisirs et sociabilités en ligne. Dans chacun de ces domaines, des données ou informations sont utilisées, produites, stockées, recherchées et mobilisées dans des démarches et des façons de faire très diverses. Selon les individus rencontrés, l'expression « données personnelles » suscitait plus ou moins de réactions, voire d'incompréhensions. En se présentant sous l'appellation « MesInfos », l'enquête permettait ainsi une pluralité de perspectives, et de façons, pour des individus, de se sentir intéressé, concerné ou tout simplement curieux.

Les seconde moitié de l'année 2013 et l'année 2014 ont été consacrées à la mise en oeuvre d'un suivi de l'expérimentation par une équipe de recherche articulant un important volet quantitatif (questionnaire en 3 vagues) à un volet qualitatif plus modeste, mais complémentaire :

- une dizaine d'entretiens, dont 8 approfondis et intégralement retranscrits, ont été réalisés entre mai et juin 2014, au terme de l'expérimentation
 - 4 des 8 entretiens approfondis, qui ont pu se dérouler en face-à-face chez les interviewés, à Paris (1) et en région parisienne (3), ont fourni des occasions d'observation : des ordinateurs, des comptes utilisés durant l'expérimentation, et plus généralement de l'écologie de « leurs infos, leurs données et documents » au domicile. Ces entretiens ont duré entre une et deux heures
 - Les 4 autres, réalisés par téléphone avec des personnes résidant en province (ville moyenne, petite ville et en campagne), ont
- une observation du *focus group* organisé avec une dizaine « testeurs » a été réalisée en mai 2014 dans les locaux de la Fing, qui visait à faire commenter les applications à des panélistes particulièrement impliqués et/ou disserts, et à obtenir des retours sur leurs premières impressions d'usage de quelques services et/ou scénarios de services.

Par delà la confiance : élargir le répertoire des usages et des interprétations des données personnelles

Les personnes enquêtées dans une telle enquête ne peuvent bien sûr prétendre à aucune représentativité statistique, mais ont été choisies de façon à tester la variabilité potentielle de leurs pratiques en fonction de plusieurs critères, tels que l'âge, le sexe et le lieu de résidence. Une attention a été portée pour saisir les significations des activités et des pratiques des individus en vue de compléter le travail d'analyse des usages aux prises avec la plateforme accompli par l'intermédiaire du questionnaire. On a également veillé à ne sélectionner que quelques profils (2) exerçant une activité professionnelle proche des mondes informatiques et du web, qui se trouvent effectivement parmi les plus impliqués dans le cours des entretiens. Les contraintes de l'enquête liée à l'expérimentation étant fortes (taux d'abandon inéluctable, contrainte représentée par un entretien, etc.), le nombre d'entretien demeure très modeste, et les résultats restent à interpréter avec prudence : surtout, ils ne permettent que peu, voire pas, de généralisation. Ils font toutefois émerger des façons de produire, de gérer et de vivre avec « ses données personnelles » particulièrement riches, discutées et sujettes à évolution, dans un contexte élargi de questionnement public renouvelé pour les questions d'informatique et de protection de la vie privée.

Des enquêtés rétribués, des clients réflexifs

Enfin, une dernière précision importante pour les partenaires de l'enquête tient aux modalités de constitution du panel par la société en charge de la communauté d'expérimentation. **Comme c'est souvent le cas dans de telles études, le panel constitué l'a été au prisme d'un intérêt pour les attitudes et les pratiques de consommation des panélistes. Ceux-ci, qui ont volontairement participé à l'expérimentation, ont pu recevoir des incitations, et notamment un « dédommagement » financier pour la passation d'un entretien.** Pour la moitié des personnes interviewées, répondre à des enquêtes en ligne est une pratique régulière ; expliciter ses modes de consommation également. De ce point de vue, s'il faut souligner l'intérêt de ces habitués pour une expérimentation dont ils ont souvent salué l'originalité, il faut aussi insister sur le rapport soutenu que l'essentiel d'entre eux ont l'habitude d'entretenir avec des sondages marketing ou des marques. **On peut tendanciellement faire l'hypothèse que ceux-ci ont plutôt, en règle générale, une certaine confiance dans les entreprises dont ils sont clients. Consommateurs avertis voire impliqués, leur expérience (générale, et préalable à l'expérimentation) les conduits à considérer qu'il n'y a pas de relation de service possible sans, d'une façon ou d'une autre, délivrer nombre de « données personnelles » et laisser la possibilité d'effectuer un travail de réanalyse avec la promesse - l'espoir - d'un service meilleur.** Les interrogations, qui peuvent être vives, portent alors surtout sur la délégation à des tiers des informations fournies et/ou collectées par les entreprises.

2. La confrontation à ses données personnelles : choc ou non-événement ?

On le sait, cela rejoint d'une autre façon, des résultats connus... Un premier résultat fort, le "paradoxe de la vie privée" : un aspect décisif pour les individus ("oui, il y a un risque"), qui semble souvent entrer en contradiction avec la foule de leurs explorations, des risques qu'ils semblent prendre : dans la pratique, on s'en ficherait. Les travaux pointant que "la fin de la vie privée" n'a pas lieu - même qd on étudie les pratiques des plus jeunes sur Facebook - cf. Antonio Casilli ; les travaux sur la

privacy by design de B. Rey et E. Kessous. C'est le premier mérite de cette expérimentation : produire un cadre original d'expériences durant plusieurs mois, qui fournit des appuis pour interroger et interviewer une dizaine d'individus sur leurs pratiques, leur étonnement ou leur désintérêt (partiel) :

- "Recrutés" parmi un panel d'enquêtés dont un des intérêts est de n'être pas complètement rétif à un usage marchand de données, en tant que consommateurs : "cartes de fidélité"
- Marc¹ résume ainsi de quoi il était question pour lui : "(...) enfin les entreprises vont mettre à disposition les données personnelles qu'ils ont sur nous et en gros, on allait voir des choses... donc au début c'était, c'était un concept que je n'avais jamais vu, justement, les données personnelles tout le monde en parle et je me suis dit « tiens, voilà, on va voir ce que ça peut donner, est-ce que je peux voir justement ce que font les entreprises ? » ; puis "des choses sur moi, mes manières"...

Il n'y a pas eu de "choc" face au "don" des données parmi les quelques personnes interrogées, pas du moins devant la possible "énormité" des données fournies ; mais il y a eu, dans un second temps, trois types de soubresauts, de questionnements renouvelés pour la plupart des enquêtés :

- Une expérience délimitée : l'usage des adresses de courrier électronique jetables proposées par le service de Priwovny : permettait de prendre concrètement la mesure de l'ampleur des pratiques de cession et revente de données personnelles à des tiers, par des sites web fréquentés (en premier lieu, le cas de l'adresse mail).
- Par l'apprentissage tout au long de l'expérimentation :
 - découverte de l'analyse de son caddie de courses alimentaires par types de produits ;
 - de la mise en relation potentielle entre ce qu'il y a dans un placard de la cuisine, et ce qu'il faut racheter au supermarché, etc. : **de nouvelles questions émergent pour les individus en regard de leur histoire personnelle** (expérience centrale du spam et de la publicité contextuelle et ciblée depuis 5 à 10 ans au moins)
- Un troisième motif de soubresaut sur ces thématiques, plus général, lié à la diffusion médiatique d'enjeux liés à la surveillance, au stockage et à la réutilisation de données de communications privées sur le web (NSA, etc.) : qui est mis en lien, de façon pas nécessairement angoissée, avec l'expérimentation. Ce troisième motif souligne l'importance, lorsqu'on cherche à approcher la façon dont les individus considèrent les enjeux de *privacy*, à considérer celle-ci « en contexte », et les évolutions socio-historiques des significations de la privacité (Nissenbaum, 2010).

¹ Les prénoms des enquêtés ont été modifiés.

3. « Mes » données, ou « leurs » données ? Confrontation entre données sur soi et traces de soi produites par une entreprise, et rôles pour les intermédiaires ?

- **Résultat intermédiaire** : approfondissement via l'expérimentation de ce que veut dire ou pas "données personnelles" (de courses ?) ? Comment en parler, ne pas bloquer les individus interrogés ?
- **Données...** : de quelle **personne(ille)** parle-t-on ? données d'**identité** / données administratives / données consommateur / utilisateur (internaute)...
- **...ou Traces ? Traces d'activités et données transactionnelles** = médiation : moi, autrui, un individu, une institution : remise en cause que tout est personnel et possède cette délimitation "personnelle" : on donne/dévoile un peu de soi dans tout échange social - ce qui en ressort n'est pas irrémédiablement personnel (ex : un log de connexion web) ; les données, ce n'est pas que de la vie privée - c'est de la connaissance (relationnelle)
- Expérience étonnante car d'habitude : Les "données" ne sont pas "données" : terme impropre : elles sont produites, ce sont des productions ; il faut les réclamer, on ne les a pas si souvent, si tant est qu'on les veuille... et elles ont longtemps été coûteuses et pas données ; cf. Latour : des "obtenues" !
 - Pour l'individu, il faut les archiver, les classer, pouvoir y rechercher telle ou telle information, tel élément : leurs routines se situent plus à l'échelle des documents qu'à l'échelle de données (bien qu'ils soient nombreux à avoir essayé Excel)
 - **Le travail des données chez soi**, au domicile (slide 2) : quasiment toutes les personnes vérifient une partie des données qui leur sont transmises, telles que leurs données bancaires :
 - *"L'intérêt c'était effectivement de voir qu'on pouvait avoir un outil qui permettait de regrouper les données et quitte à ce que nous, on voit uniquement la partie visuelle des données qui est celle qui s'affiche sur l'écran mais qu'en arrière-plan le système sache aller piocher dans différents endroits. C'est le regroupement des données, parce qu'effectivement au quotidien on en a un peu partout, disséminé"*
 - Parfois, pointe une critique de l'injonction à gérer soi-même ses données, à la seule responsabilisation du consommateur
 - Ce sont les données de l'entreprise... pour une bonne partie ! (traces // scoring). Ce que leur banque sait sur eux vs. ce qu'elle pense d'eux... et peut leur accorder) : les données et leurs interprétations (entre confiance et suspicion - "je ne sais pas ce qu'ils peuvent savoir avec ça... C'est cela qu'ils devraient pouvoir me dire")

4. Petite esquisse de typologie des types et cercles de « données »

À l'issue de l'enquête, le protocole expérimental permet de mettre en évidence la tension qu'il y a, pour les individus, à considérer selon les situations la nature plus ou

moins... personnelle des « données personnelles ». Celles-ci ne sont pas toujours seulement personnelles, ou même simplement ajustées à l'individu. Les entretiens font ressortir que ces informations sont entrelacées dans plusieurs échelles, de la plus ajustée à l'individu (le désignant, permettant de l'identifier) à une échelle plus attentive aux différents rôles sociaux des individus, mais aussi à une échelle plus collective (les données qui font sens pour le couple, les enfants, la famille, un groupe, etc.) :

- Comment **catégoriser** "ses" données "personnelles" ? (vraies données personnelles, et encore : ses photos !). Son prénom, et son nom, sa date de naissance si associée aux premiers éléments. Son numéro de sécurité sociale. Selon les cas, un numéro de carte bleue (souvent aussi lié au couple, au « ménage »)
- Cozycloud donne l'idée d'obtenir des données pour "comparer" : consommation électrique par que pour son foyer, ou une moyenne générale pour la France, mais par rapport à son quartier, et plus encore, à son immeuble (ie. prendre une décision sur l'isolation, par ex.) ! La donnée fournie par l'entreprise, le service, trouve son sens à travers la comparaison avec autrui, ou des collectifs.
- **De MesInfos à... "NosInfos"** (ex. de la classification des dépenses de courses) : consommation, soi... couple, enfants, proches, l'échelle du foyer : les données perso de conso existent-elles si fréquemment ? Cf. aussi histoire de l'administration électronique (cf. risques sécurité et certificat électronique... à login/mdp) :
 - "MesInfos", ce serait sur **le mobile**, beaucoup plus que sur **l'ordinateur plus ou moins partagé voire tout à fait collectif**, et seul objet véritablement individualisé pour tous les interviewés ;
 - Proposer plusieurs échelles d'analyse : soi ; couple/foyer / avec enfants / cercles de sociabilité à intérêts potentiellement communs : voisins, etc.
 - Enfin, idéalement, proposer des données commerciales, mais pas que... avec notamment des données administratives d'une part, voire des données de santé ou d'activités physiques/corporelles d'autre part, notamment pour susciter des recherches de corrélation et des "découvertes" plus vastes

L'expérimentation conduit une bonne partie des individus à étendre leurs questionnements et développer leur analyse en déconstruisant d'abord la catégorie "personnelle". Des données qui ne sont pas à soi, des données qui disent peu de vous-mêmes, ou beaucoup, ou trop, pour vous, pour autrui, etc. :

- Attente de symétrisation de la relation avec des organisations
- Évoquant des données chiffrées telles que le poids, la taille et plus encore, les données de technologies d'auto-mesure du QS : (Marie) "Là, ça c'est quand même très personnel. C'est un autre domaine des données personnelles qui sont vraiment nos données viscérales."

- **Au final, ils et elles questionnent, entre curiosité et scepticisme, l'idée d'un mode de connaissance de soi par l'infra-quotidien (idée que les choses banales nous décriraient le mieux) ; connaissance descriptive objectivante, très comportementale de soi, complémentaire d'une forme plus introspective.**

5. Premiers éléments de conclusion

Les enquêtés se caractérisent plutôt par le fait qu'ils se sont résolus, au fil du temps, à accorder une certaine confiance aux entreprises dont ils sont clients. Nous n'avons pas échangé, par exemple, avec des interviewés qui auraient témoigné de leur défiance systématique. S'ils peuvent se montrer surpris lorsqu'ils constatent - souvent en surface - la richesse des données parfois réunies (par l'opérateur téléphonique notamment) et utilisées par ces entreprises, leur confiance se traduit par l'espoir que ces informations sont utiles pour fournir un service plus ajusté. Dans le cas des achats au supermarché, néanmoins, si les interviewés reconnaissent l'intérêt du codage de leurs dépenses pour mieux gérer son budget, et laissent entendre de nouvelles attentes survenues pendant l'expérimentation *MesInfos*, celles-ci s'accompagnent d'un questionnement : dans quelle mesure ces données sont-elles travaillées pour permettre d'articuler les objectifs des entreprises d'une part, et des consommateurs d'autre part - lesquels peuvent selon les cas converger... et/ou diverger ? Avec quel équilibre ? Cette interrogation n'est pas synonyme, pour ces quelques testeurs du dispositif *MesInfos*, d'un point d'arrêt dans leur réflexion ou du surgissement exclusif de nouvelles craintes, mais plutôt du souhait que ce type d'expérimentations parvienne, pour partie au moins, à maturité.

EN SAVOIR PLUS

TOUT SAVOIR SUR LE PROJET ET LE CONCOURS MESINFOS

Le Site - <http://mesinfos.fing.org>

Le Blog - <http://mesinfos.fing.org/blog>

CONTACT

Marine Albarède - malbarede@fing.org

Renaud Francou - rfrancou@fing.org

Sarah Medjek - smedjek@fing.org

PARTENAIRES



PARTENAIRES ASSOCIÉS

