

Enquête usagers, découvrabilité des données

Céline Goffette

22 octobre 2020

Table of Contents

Profil des répondants à l'enquête usagers.....	2
Activité professionnelle.....	2
Compétences en manipulation de données.....	3
Questions liées à la découvrabilité des données.....	8
“Déterminants” de la découvrabilité.....	11
Mentions au coronavirus dans les questions ouvertes.....	14
Sélection de verbatims liés à la découvrabilité.....	15
Propositions de slides pour la présentation des résultats	17
Slide 1 : L'enquête auprès des usagers de data.gouv.fr.....	17
Slide 2 : Plus de gens du privé que du public, et des activités professionnelles variées.....	18
Slide 3 : Des usagers pas tous rompus à la manipulation de données.....	19
Slide 4 : La moitié des répondants pas ou peu formés à la manipulation des données	20
Slide 5 : La manipulation des données fait partie de l'activité professionnelle de la moitié des répondants	21
Slide 6 : Les répondants arrivent en majorité à trouver ce qu'ils cherchent.....	23
Slide 7 : Quand les recherches se soldent par un échec.....	24
Slide 8 : Un accès principal aux données via un moteur de recherche externe.....	25
Slide 9 : Qui trouve, qui ne trouve pas son jeu de données ?	26
Slide 10 et suivantes : Retours d'utilisateurs sur la découvrabilité des données.....	27

Profil des répondants à l'enquête usagers

Enquête auprès des usagers de la plateforme data.gouv.fr
<https://framaforms.org/enquete-datagouvfr-usagers-1590438305>
N = 902 répondants

Activité professionnelle

“Vous vous définissez comme ?” (plusieurs réponses possibles)

- agent-e de la fonction publique (*public*)
- salarié-e du privé (*prive*)
- chercheur ou chercheuse (*chercheur*)
- enseignant-e (*enseignant*)
- statisticien-ne, data scientist, data analyst (*statistiques*)
- développeuse / développeur informatique (*developpement*)
- journaliste (*journaliste*)
- militant-e (*militant*)
- étudiant-e (*étudiant*)
- autre (*autreprofession*)

Caractéristique	N = 902 ¹
public	194 (22%)
prive	263 (29%)
chercheur	70 (7,8%)
enseignant	56 (6,2%)
statistiques	125 (14%)
developpement	147 (16%)
journaliste	8 (0,9%)
militant	50 (5,5%)
etudiant	84 (9,3%)
autreprofession	285 (32%)

¹Statistique présentée: n (%)

A noter : 227 individus ayant répondu “autre” (*autreprofession*) ont coché uniquement cette modalité. On n’a donc aucune information sur l’activité professionnelle pour un quart des répondants.

Malgré la possibilité de cocher plusieurs réponses, une majorité des répondants (573, soit 63%) n’a coché qu’une seule modalité. Les modalités les plus fréquemment associées sont “salarié-e du privé, statisticien-ne, data scientist, data analyst”, mais cela ne concerne que 24 individus soit 2,7% (résultat non présenté ici, mais possibilité de visualiser le tableau depuis le fichier .Rmd [code en commentaire au-dessus de ce paragraphe]).

Compétences en manipulation de données

“Vous estimez que vos compétences techniques en manipulation de données sont ?”

- Fortes : vous manipulez souvent des données ou des bases de données, vous maîtrisez un ou plusieurs langage(s) de programmation, vous avez déjà développé un outil basé sur des données... (*Fortes*)
- Moyennes : vous savez manipuler plusieurs formats de données, faire quelques calculs ou agrégations, visualiser des données sans vous tromper (*Moyennes*)
- Minimales : vous avez déjà vu des jeux de données mais vous n’avez pas su les exploiter (*Minimales*)
- Non renseigné (*Manquant*)

Caractéristique	N = 902 ¹
V3_rec	
Minimales	187 (21%)
Moyennes	405 (46%)
Fortes	285 (32%)
Manquant	25

¹Statistique présentée: n (%)

“Vous avez des compétences en manipulation de données. En particulier, vous savez”

- utiliser un ou plusieurs langage(s) de programmation (*V4*)
- manipuler des bases de données (SQL, MySQL)... (*V5*)
- manipuler des données textuelles (*V6*)
- manipuler des données géographiques (*V7*)
- créer des visualisation de données avec un logiciel (Excel, LibreOffice, Google Docs...) (*V8*)
- créer des visualisation de données dans un langage de programmation (d3.js, ggplot...) (*V9*)
- utiliser un outil de BI (Tableau Software, PowerBI, OpenDataSoft...) (*V10*)

Caractéristique	Minimales, N = 187 ¹	Moyennes, N = 405 ¹	Fortes, N = 285 ¹
V4			
pas du tout	141 (82%)	141 (36%)	8 (2,9%)
un peu	23 (13%)	160 (41%)	23 (8,2%)
bien	4 (2,3%)	68 (18%)	90 (32%)
très bien	3 (1,8%)	18 (4,7%)	158 (57%)
Manquant	16	18	6

V5

Caractéristique	Minimes, N = 187 ¹	Moyennes, N = 405 ¹	Fortes, N = 285 ¹
pas du tout	146 (86%)	140 (36%)	11 (3,9%)
un peu	19 (11%)	154 (40%)	22 (7,8%)
bien	5 (2,9%)	73 (19%)	91 (32%)
très bien	0 (0%)	18 (4,7%)	157 (56%)
Manquant	17	20	4
V6			
pas du tout	87 (52%)	26 (6,8%)	5 (1,8%)
un peu	56 (33%)	129 (34%)	42 (15%)
bien	21 (12%)	169 (44%)	85 (31%)
très bien	4 (2,4%)	59 (15%)	144 (52%)
Manquant	19	22	9
V7			
pas du tout	101 (60%)	69 (18%)	19 (6,9%)
un peu	53 (32%)	143 (37%)	77 (28%)
bien	11 (6,6%)	104 (27%)	73 (26%)
très bien	2 (1,2%)	70 (18%)	107 (39%)
Manquant	20	19	9
V8			
pas du tout	68 (40%)	24 (6,1%)	11 (3,9%)
un peu	65 (38%)	83 (21%)	31 (11%)
bien	28 (17%)	173 (44%)	106 (38%)
très bien	8 (4,7%)	111 (28%)	133 (47%)
Manquant	18	14	4
V9			
pas du tout	161 (95%)	291 (75%)	72 (26%)
un peu	6 (3,6%)	62 (16%)	78 (28%)
bien	2 (1,2%)	22 (5,7%)	54 (20%)
très bien	0 (0%)	12 (3,1%)	71 (26%)
Manquant	18	18	10
V10			
pas du tout	153 (92%)	223 (59%)	98 (36%)
un peu	13 (7,8%)	97 (25%)	71 (26%)
bien	0 (0%)	48 (13%)	53 (20%)
très bien	1 (0,6%)	13 (3,4%)	49 (18%)
Manquant	20	24	14

¹Statistique présentée: n (%)

Les répondants jugeant leurs compétences minimales savent rarement utiliser un langage de programmation. Mais ils ont fréquemment des connaissances au minimum basiques en manipulation de données textuelles (50%) ou géographiques (40%), ou savent créer des visualisations de données avec un logiciel (60%).

“Avez-vous été formé-e à manipuler des données ?”

- Oui, c’est au coeur de ma formation (mathématiques, économie, statistique, data science, big data...) (*Formé à la MD (central)*)
- Oui, c’est présent dans ma formation, mais pas de manière centrale (informatique, gestion, sciences sociales...) (*Formé à la MD (présent)*)
- Non, pas vraiment, ou très peu (*Pas ou peu formé à la MD*)
- Non renseigné (*Manquant*)

Caractéristique	N = 902 ¹
V11_rec	
Peu formé	433 (49%)
Présent	294 (33%)
Central	159 (18%)
Manquant	16

¹Statistique présentée: n (%)

Près de 8 répondants sur 10 estiment avoir des compétences moyennes ou fortes en matière de manipulation de données. Pourtant, la moitié des répondants n’a pas ou peu été formée à la manipulation de données.

“La manipulation de données fait-elle partie de votre activité professionnelle ?”

- Oui
- Non
- Non renseigné (*Manquant*)

Caractéristique	N = 902 ¹
V12_rec	
Oui	474 (53%)
Non	412 (47%)
Manquant	16

¹Statistique présentée: n (%)

Utilisation dans le cadre professionnel x formation et compétences en manipulation de données

Caractéristique	Oui, N = 474 ¹	Non, N = 412 ¹
V11_rec		
Peu formé	135 (28%)	295 (72%)
Présent	202 (43%)	92 (23%)
Central	137 (29%)	20 (4,9%)
Manquant	0	5
V3_rec		
Minimes	21 (4,5%)	165 (41%)
Moyennes	216 (46%)	186 (46%)
Fortes	232 (49%)	52 (13%)
Manquant	5	9

¹Statistique présentée: n (%)

La manipulation de données fait partie de l'activité professionnelle de la moitié des répondants. Mais parmi eux, 28% n'ont pas ou peu été formés à la manipulation de données, et 46% jugent leurs compétences en la matière moyennes.

“Combien de fois êtes vous déjà allé-e sur l'un ou l'autre de ces sites (data.gouv.fr, transport.data.gouv.fr, adresse.data.gouv.fr, etc.), au total, au cours des 12 derniers mois ?”

- 1-2 fois
- 3-10 fois
- 11-30 fois
- Plus de 30 fois
- Non renseigné (*Manquant*)

Caractéristique	N = 902 ¹
V15	
1-2 fois	234 (27%)
3-10 fois	285 (32%)
11-30 fois	139 (16%)
Plus de 30 fois	223 (25%)
Manquant	21

¹Statistique présentée: n (%)

Fréquence d'usage x formation et compétences en manipulation de données

Caractéristique	1-2 fois, N = 234 ¹	3-10 fois, N = 285 ¹	11-30 fois, N = 139 ¹	Plus de 30 fois, N = 223 ¹
V11_rec				
Peu formé	151 (66%)	140 (49%)	49 (36%)	81 (36%)
Présent	53 (23%)	103 (36%)	57 (42%)	79 (36%)
Central	24 (11%)	42 (15%)	31 (23%)	62 (28%)
Manquant	6	0	2	1
V3_rec				
Minimes	96 (42%)	45 (16%)	18 (13%)	20 (9,0%)
Moyennes	95 (42%)	149 (53%)	66 (49%)	90 (41%)
Fortes	36 (16%)	87 (31%)	50 (37%)	111 (50%)
Manquant	7	4	5	2

¹Statistique présentée: n (%)

Un quart des répondants déclare avoir visité l'un ou l'autre des sites mentionnés (data.gouv.fr, transport.data.gouv.fr, adresse.data.gouv.fr, etc.) plus de 30 fois au cours des 12 derniers mois. Parmi ces utilisateurs fréquents, un tiers est peu ou pas formé à la manipulation de données, et un sur dix estime que ses compétences en la matière sont minimales.

Questions liées à la découvrabilité des données

“Vous avez parcouru le portail data.gouv.fr. Avez-vous trouvé ce que vous cherchiez ?”

- Oui
- Non
- Je ne cherchais rien de particulier
- Non renseigné (*Manquant*)

Caractéristique	N = 902 ¹
V19_rec	
Oui	472 (53%)
Non	243 (28%)
Rien	168 (19%)
Manquant	19

¹Statistique présentée: n (%)

Interprétation délicate des réponses à cette question, qui aurait demandé à être précisée.

Par exemple : “avez-vous trouvé ce que vous cherchiez *lors de votre dernière visite ?*”.

La question suivante portait sur les raisons expliquant pourquoi la recherche avait échoué. Les répondants ont souvent donné plusieurs raisons (alors qu’il n’était pas précisé que plusieurs réponses étaient possibles, mais le questionnaire le permettait), ce qui peut laisser penser qu’ils ont listé tous les problèmes auxquels ils ont déjà été confrontés en utilisant data.gouv.fr. Ils ont également indiqué plusieurs voies d’accès aux données (alors qu’il n’était également pas précisé que plusieurs réponses étaient possibles).

La moitié des répondants indiquent avoir trouvé ce qu’ils cherchaient sur data.gouv.fr. 19 % ne cherchaient rien. 28% des répondants indiquent que leur recherche s’est soldée par un échec.

“Pourquoi ?” (plusieurs réponses possibles, même si non explicitement précisé dans le questionnaire)

- Le jeu de données n’existait pas ou je n’ai pas réussi à le retrouver via le moteur de recherche, les catégories (*pastrouve*)
- Le jeu de données n’existait plus : une page existait avec un lien mort (*disparu*)
- Le jeu de données n’avait pas un format exploitable pour moi (*format*)
- Le jeu de données était mal documenté, je n’ai pas su comment l’utiliser (*documentation*)
- Le jeu de données n’avait pas la bonne granularité, les bonnes métadonnées (trop détaillé, trop agrégé, pas assez de variables...) (*granularite*)
- Je n’avais pas confiance dans les données proposées (*confiance*)
- Autre (*autreraison*)

Caractéristique	N = 371 ¹
pastrouve	157 (42%)
disparu	50 (13%)
format	67 (18%)
documentation	75 (20%)
granularite	75 (20%)
confiance	44 (12%)
autreraison	105 (28%)

¹Statistique présentée: n (%)

La question “Pourquoi ?” n’était pas filtrée, alors qu’on suppose qu’elle était implicitement dédiée aux individus qui avaient répondu “non” à la question précédente. Du fait de l’absence de filtre, certains des individus ayant répondu “oui” ou “je ne cherchais rien de particulier” ont également répondu à cette question. Au total, 371 individus ont répondu à la question (alors que seulement 243 avaient répondu “non” à la question précédente). On a décidé de garder toutes les réponses, considérant qu’elles doivent probablement être vues comme une liste de tous les problèmes déjà rencontrés par les utilisateurs de data.gouv.fr. Ceux qui ont répondu “autre” précisent le plus souvent que c’est parce que les données étaient datées, n’étaient pas actualisées (exploitation ouvert **“Si vous avez coché”autre”, pouvez-vous préciser ?**) : “données cherchées trop anciennes”, “données non actualisées”, données obsolètes”, “jeu de données pas à jour”, “trop ancien”.

Caractéristique	Oui, N = 76 ¹	Non, N = 230 ¹	Rien, N = 64 ¹
pastrouve	27 (36%)	120 (52%)	9 (14%)
disparu	17 (22%)	30 (13%)	2 (3,1%)
format	23 (30%)	37 (16%)	6 (9,4%)
documentation	14 (18%)	47 (20%)	13 (20%)
granularite	23 (30%)	41 (18%)	11 (17%)
confiance	11 (14%)	16 (7,0%)	17 (27%)
autreraison	23 (30%)	59 (26%)	23 (36%)

¹Statistique présentée: n (%)

“Par quel chemin êtes vous arrivé à ce jeu de données ?” (plusieurs réponses possibles, même si non explicitement précisé dans le questionnaire)

- Accès direct via moteur de recherche (Google, Bing, DuckDuckGo...) (*moteur*)
- Article de blog (*blog*)
- Catégories thématiques (*categories*)
- Moteur de recherche interne data.gouv.fr (*interne*)
- Autre (*autreaces*)

Caractéristique	N = 840 ¹
moteur	495 (59%)
blog	93 (11%)
categories	76 (9,0%)
interne	331 (39%)
autreaces	91 (11%)

¹Statistique présentée: n (%)

Ici également la question aurait demandé à être précisée. On ne sait pas bien à quoi “ce jeu de données” fait référence, et à qui s’adressait cette question. Probablement à ceux ayant répondu “oui” à la question “Vous avez parcouru le portail data.gouv.fr. Avez-vous trouvé ce que vous cherchiez ?”.

Du fait de l’absence de filtre, une part importante des individus ayant répondu “non” ou “je ne cherchais rien de particulier” a également répondu. Au total, 840 individus ont répondu à la question. On a décidé de garder toutes les réponses, considérant qu’elles doivent probablement être vues comme les différentes voies d’accès aux données auxquelles ont recours les utilisateurs de data.gouv.fr.

39% des répondants disent avoir trouvé le jeu de données via le moteur interne de data.gouv.fr, et 59% via un moteur de recherche externe. Ceux qui ont répondu “autre” précisent le plus souvent qu’ils sont arrivés au jeu de données via un autre site (exploitation ouvert **“Si vous avez répondu”autre”, pouvez-vous préciser ?”**).

Caractéristique	Oui, N = 457 ¹	Non, N = 236 ¹	Rien, N = 146 ¹
moteur	289 (63%)	127 (54%)	79 (54%)
blog	49 (11%)	21 (8,9%)	23 (16%)
categories	48 (11%)	17 (7,2%)	10 (6,8%)
interne	199 (44%)	94 (40%)	37 (25%)
autreaces	49 (11%)	22 (9,3%)	19 (13%)

¹Statistique présentée: n (%)

“Déterminants” de la découvrabilité

“Vous estimez que vos compétences techniques en manipulation de données sont ?”

Caractéristique	Minimes, N = 187 ¹	Moyennes, N = 405 ¹	Fortes, N = 285 ¹
V19_rec			
Oui	69 (38%)	230 (57%)	164 (58%)
Non	60 (33%)	106 (26%)	76 (27%)
Rien	55 (30%)	64 (16%)	43 (15%)
Manquant	3	5	2
pastrouve	29 (16%)	71 (18%)	56 (20%)
disparu	7 (3,7%)	21 (5,2%)	22 (7,7%)
format	14 (7,5%)	28 (6,9%)	24 (8,4%)
documentation	17 (9,1%)	32 (7,9%)	26 (9,1%)
granularite	7 (3,7%)	28 (6,9%)	40 (14%)
confiance	19 (10%)	7 (1,7%)	17 (6,0%)
autreraison	28 (15%)	41 (10%)	32 (11%)
moteur	97 (52%)	227 (56%)	162 (57%)
blog	16 (8,6%)	41 (10%)	35 (12%)
categories	9 (4,8%)	34 (8,4%)	31 (11%)
interne	38 (20%)	155 (38%)	134 (47%)
autreaces	25 (13%)	42 (10%)	22 (7,7%)

¹Statistique présentée: n (%)

Les répondants qui estiment avoir des compétences minimales en manipulation de données trouvent un peu moins ce qu'ils cherchent (38% déclarent avoir trouvé, contre 57% pour ceux qui ont des compétences moyennes et 58% pour ceux qui ont des compétences fortes). Ils déclarent aussi plus souvent qu'ils ne recherchaient rien de particulier. Ils utilisent un peu moins le moteur de recherche interne de data.gouv.fr que les autres. En ce qui concerne les problèmes rencontrés, on ne note pas beaucoup de différences en fonction du niveau de compétences. Ceux qui ont des compétences minimales déclarent un peu plus fréquemment ne pas avoir confiance dans les données, et ceux qui ont des compétences fortes citent des problèmes liés à la granularité.

“Avez-vous été formé-e à manipuler des données ?”

Caractéristique	Peu formé, N = 433 ¹	Présent, N = 294 ¹	Central, N = 159 ¹
V19_rec			
Oui	208 (49%)	175 (60%)	87 (55%)
Non	123 (29%)	76 (26%)	43 (27%)
Rien	93 (22%)	43 (15%)	27 (17%)
Manquant	9	0	2
pastrouve	76 (18%)	56 (19%)	25 (16%)
disparu	17 (3,9%)	24 (8,2%)	9 (5,7%)
format	32 (7,4%)	24 (8,2%)	11 (6,9%)
documentation	33 (7,6%)	26 (8,8%)	16 (10%)
granularite	23 (5,3%)	31 (11%)	20 (13%)
confiance	23 (5,3%)	10 (3,4%)	10 (6,3%)
autreraison	49 (11%)	33 (11%)	22 (14%)
moteur	236 (55%)	177 (60%)	79 (50%)
blog	36 (8,3%)	36 (12%)	17 (11%)
categories	28 (6,5%)	35 (12%)	13 (8,2%)
interne	118 (27%)	127 (43%)	84 (53%)
autreaces	51 (12%)	26 (8,8%)	14 (8,8%)

¹Statistique présentée: n (%)

Les répondants qui sont peu ou pas formés à la manipulation de données trouvent un peu moins ce qu'ils cherchent (49% déclarent avoir trouvé, contre 60% et 55%). L'utilisation du moteur de recherche interne de data.gouv.fr croît avec le niveau de compétences déclaré.

“La manipulation de données fait-elle partie de votre activité professionnelle ?”

Caractéristique	Oui, N = 474 ¹	Non, N = 412 ¹
V19_rec		
Oui	283 (60%)	185 (46%)
Non	116 (25%)	127 (31%)
Rien	72 (15%)	92 (23%)
Manquant	3	8
pastrouve	91 (19%)	66 (16%)
disparu	33 (7,0%)	17 (4,1%)
format	39 (8,2%)	28 (6,8%)
documentation	40 (8,4%)	35 (8,5%)
granularite	60 (13%)	15 (3,6%)
confiance	21 (4,4%)	21 (5,1%)
autreraison	49 (10%)	55 (13%)
moteur	278 (59%)	211 (51%)
blog	50 (11%)	42 (10%)
categories	54 (11%)	22 (5,3%)
interne	212 (45%)	117 (28%)
autreaces	42 (8,9%)	47 (11%)

¹Statistique présentée: n (%)

Les répondants dont la manipulation de données fait partie de l'activité professionnelle trouvent plus fréquemment ce qu'ils cherchent. Ils utilisent également plus le moteur de recherche interne de data.gouv.fr.

Mentions au coronavirus dans les questions ouvertes

Le questionnaire en ligne a été ouvert à partir du 26 juin 2020. Plusieurs visites sur le site data.gouv.fr ont pu être motivées par la crise sanitaire liée au coronavirus. Les termes “covid”, “Covid”, “COVID”, “coronavirus” et “cluster” ont été recherchés dans les questions ouvertes pour déterminer l’ampleur de cet “effet coronavirus” sur les visites du site.

Les questions ouvertes sont les suivantes :

- [Vous avez parcouru le portail data.gouv.fr. Avez-vous trouvé ce que vous cherchiez ? Pourquoi ?] **Si vous avez coché “autre”, pouvez-vous préciser ?** (*coronavirusV21*)
- **Souhaitez-vous ajouter des précisions quant à vos réponses précédentes ?** (*coronavirusV22*)
- [Par quel chemin êtes vous arrivé à ce jeu de données ?] **Si vous avez répondu “autre”, pouvez-vous préciser ?** (*coronavirusV24*)
- **Avez-vous des remarques, des questions, des commentaires, que vous aimeriez nous communiquer, pour améliorer data.gouv.fr, ou pour une autre raison ?** (*coronavirusV29*)
- Occurrence dans au moins un de ces questions (*coronavirus*)

Caractéristique	N = 902 ¹
coronavirusV21	6 (0,7%)
coronavirusV22	7 (0,8%)
coronavirusV24	8 (0,9%)
coronavirusV29	18 (2,0%)
coronavirus	27 (3,0%)

¹Statistique présentée: n (%)

Les termes “covid”, “Covid”, “COVID”, “coronavirus” ou “cluster” sont mentionnés par 3% des individus dans leurs réponses au questionnaire.

Sélection de verbatims liés à la découvrabilité

Avez-vous des remarques, des questions, des commentaires, que vous aimeriez nous communiquer, pour améliorer data.gouv.fr, ou pour une autre raison ?

“A l’usage, on note la présence de nombreux jeux de données portant sur les **mêmes éléments mais publiés par des producteurs différents**. Regrouper ces données et signaler l’existence de jeux de données similaires aux producteurs seraient utiles pour améliorer la qualité des données et réduire les doubles productions inutiles.”

“Ayant un usage important de données concernant des territoires spécifiques et une exploitation par analyse spatiale, **une recherche par emprise géographique permettrait de faciliter la recherche de toutes les données (éventuellement ou pas géolocalisées) de ce territoire** (si les méta-données le permettent).”

“C’est déjà une belle plateforme, très content qu’elle existe ! Idéalement, **ce serait intéressant que les données d’autres portails ressortent également dans le moteur de recherche (moissonnage)** pour simplifier les recherches de données (millefeuille territorial + pléthore de portails différents).”

“Ce serait bien que data.gouv s’adresse à une communauté large et de façon pédagogique. **il manque une dimension de clarté de documentation rapide pour comprendre le jeu de données, dictionnaire, profondeur historique...** Ce serait bien aussi que Etalab propose des **bonnes pratiques pour "normer" les données d’open data** - tjs le même nom pour les département, code com insee... a voir avec ce qui a été initié par Open data France auprès des collectivités....”

“data.gouv.fr est un site de référence et parmi les mieux faits de l’écosystème des sites publics français. Sur la forme, peu de choses à redire, si ce n’est le fait que **quand un producteur référence beaucoup de jeux de données, il n’est pas toujours facile de retrouver le jeu qui nous intéresse**. On peut aussi déplorer la documentation souvent pauvre des jeux de données (description des champs...) Enfin, **la question du maintien à jour des données et des liens est cruciale : je suis souvent tombé sur des liens morts.**”

“Dans le cas des antennes de l’ANFR (idem ailleurs ?), il manque un lien direct sur la dernière version des données. **Le nommage des fichiers n’est pas consistant**. Un système par date et/ou version homogène si possible dans tous les jeux serait pratique. Idem pour un filtre par date/période à minima pour trouver les fichiers, et mieux, directement sur les données De mémoire, trouver la description des “tables” n’est pas évident.”

“Déjà : merci infiniment pour ces données, c’est une mine d’or Pistes d’amélioration : **concernant les données géographiques, la navigation pourrait être améliorée en pouvant identifier l’ordre de grandeur : mondial, national, régional, départemental, intercommunal, communal** Idéalement avec une arborescence Je pense que bien séparer les dépôts de données (ex. BDD), et leurs utilisation (ex. une carte faite à partir de plusieurs jeux de données) est important pour la clarté de la navigation. C’est le cas aujourd’hui et je trouve ça profitable.”

“Différents producteurs produisent les mêmes données plus ou moins complètes. Cela rend difficile la recherche d’un jeu de données. Il faut ajouter un indicateur de satisfaction en plus du nombre de téléchargements. Manque de formalisme. Les jeux de données de même catégorie ont des formats différents cela rend difficile l’exploitation. Il faudrait proposer des open-format ouvert à discussion que les producteurs pourraient suivre. Il sera plus pratique de travailler avec des données de producteurs locaux. Ex portail open-data de ville.”

“English version is not at all translated (just main menu and common areas). Some international projects will need more information in english to share with other European countries.”

“il faut améliorer le catalogage des données. il faudrait pouvoir identifier facilement les données selon leur couverture géographique : 1 collectivité, 1 département, 1 région, la france. Il y a trop de donnée spécifique à un territoire, lorsqu’on mène une étude sur de grands territoires, la recherche n’est pas optimale.”

“Je rêve d’un système de notification quand un dataset est mis à jour. Typiquement sur le covid, je dois scruter les pages de Santé publique France ou du ministère de la Santé pour voir si les indicateurs de l’épidémie. Une notification (sur data.gouv, ou par mail, ou autre), simplifierait beaucoup de choses.”

“La gestion des dates (création, méta, données) est parfois erronée et au minimum perturbante. Selon la façon dont on met à jour les données, la date de mise à jour est mise à jour ... ou pas. Il faut remettre tout ça à plat et choisir un système simple, lisible et fiable. Le site est submergé de micro jeux de données qui n’en sont pas ou sont des extraits locaux d’un jeu plus général. Il faut trouver une façon de mieux mettre en valeur les vrais jeux de données. Les statistiques de vues et téléchargements sont totalement fantaisistes (certains graphiques semblent d’ailleurs avoir été partiellement retirés). Ce n’est pas sérieux pour un portail et cela empêche d’avoir des indices sur”quels sont les vrais jeux de données utiles“. **Il ne semble pas y avoir d’archivage accessible : si un producteur met à jour son jeu de données, je ne peux pas accéder aux anciennes versions. L’historisation et la non répudiation sont pourtant des conditions importantes d’un bon portail.** Cela signifie qu’on ne peut pas faire des visualisations sur des évolutions dans le temps sans faire soi-même l’effort d’archiver régulièrement. Le moteur de recherche est catastrophique. L’ordre de présentation des résultats est douteux. Les résultats fournis en auto-suggestion pendant l’autocomplétion ne sont pas cohérents (nombre, filtre, ordre) avec ceux renvoyés une fois qu’on appuie sur”Entrée“. Il y a régulièrement des problèmes de versions des pages mises dans une cache sûrement car, après avoir modifié la description d’un jeu de données, il est fréquent de retomber sur l’ancienne version quand on passe sur la page pour contrôler. C’est perturbant.”

Propositions de slides pour la présentation des résultats

Slide 1 : L'enquête auprès des usagers de data.gouv.fr

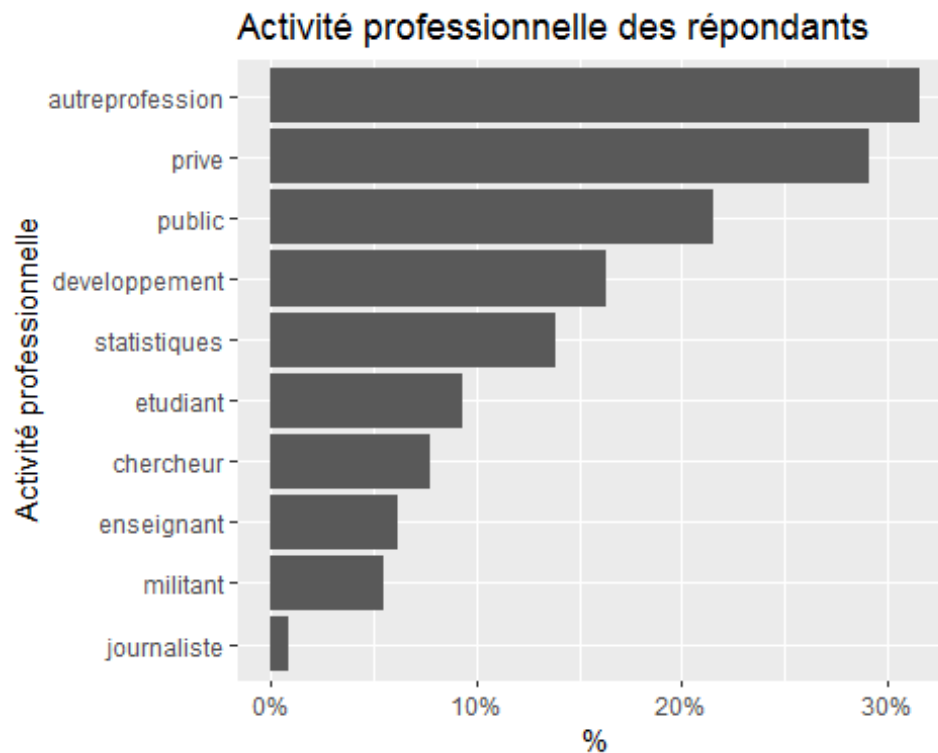
L'enquête auprès des usagers de data.gouv.fr a débuté le 26 juin 2020

902 personnes ont répondu à l'enquête

Le questionnaire comportait des questions sur la découvrabilité des données :

- Vous avez parcouru le portail data.gouv.fr. Avez-vous trouvé ce que vous cherchiez ? (oui / non / je ne cherchais rien de particulier)
- Pourquoi ? (6 raisons possibles + autre)
- Si vous avez coché "autre", pouvez-vous préciser ? (question ouverte)
- Souhaitez-vous ajouter des précisions quant à vos réponses précédentes ? (question ouverte)
- Par quel chemin êtes vous arrivé à ce jeu de données ? (4 accès possibles + autre)
- Si vous avez répondu "autre", pouvez-vous préciser ? (question ouverte)

Slide 2 : Plus de gens du privé que du public, et des activités professionnelles variées

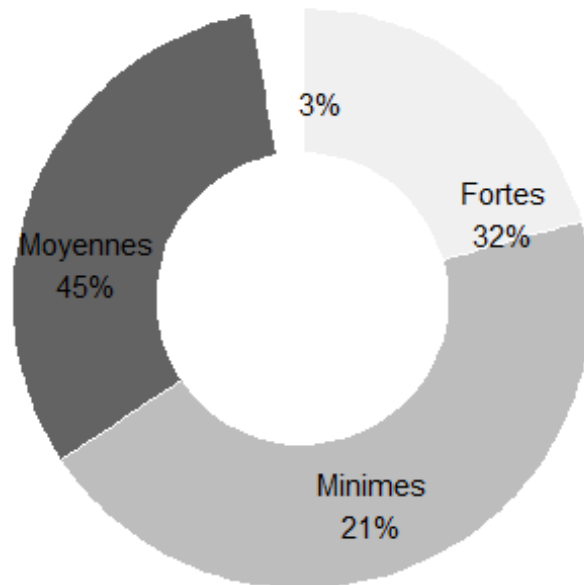


NB : plusieurs réponses étaient possibles

227 individus ayant répondu "autre" ont coché uniquement cette modalité. On n'a donc aucune information sur l'activité professionnelle pour un quart des répondants

Slide 3 : Des usagers pas tous rompus à la manipulation de données

Compétences en manipulation de données

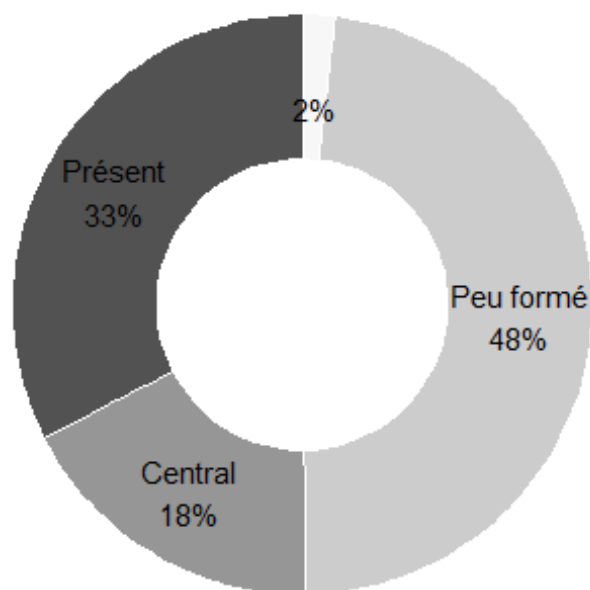


Deux répondants sur dix jugent leurs compétences en manipulation de données “minimes”

Ces usagers savent rarement utiliser un langage de programmation, mais ils ont fréquemment des connaissances au minimum basiques en manipulation de données textuelles (50%) ou géographiques (40%), ou savent créer des visualisations de données avec un logiciel (60%)

Slide 4 : La moitié des répondants pas ou peu formés à la manipulation des données

Formation à la manipulation de données



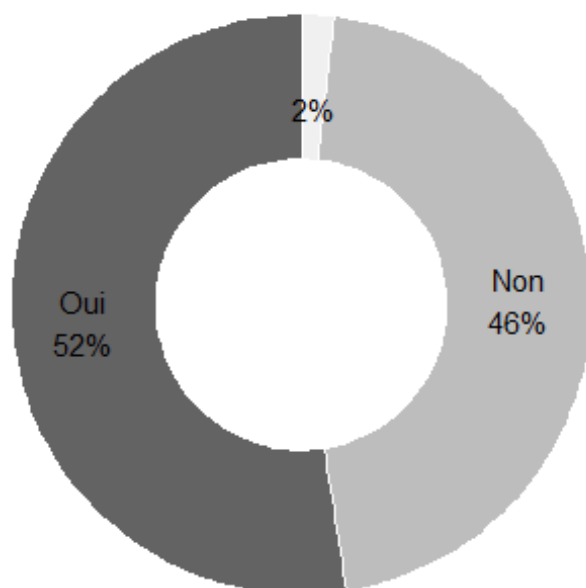
La moitié des répondants n'a pas ou peu été formée à la manipulation de données

A noter : les utilisateurs aux compétences "minimes", ou ceux qui se déclarent "pas ou peu formés" constituent une part non négligeable des usagers fréquents (= ayant visité data.gouv.fr ou un site apparenté plus de 30 fois au cours des 12 derniers mois)

Parmi les utilisateurs fréquents, un tiers est peu ou pas formé à la manipulation de données, et un sur dix estime que ses compétences en la matière sont minimales

Slide 5 : La manipulation des données fait partie de l'activité professionnelle de la moitié des répondants

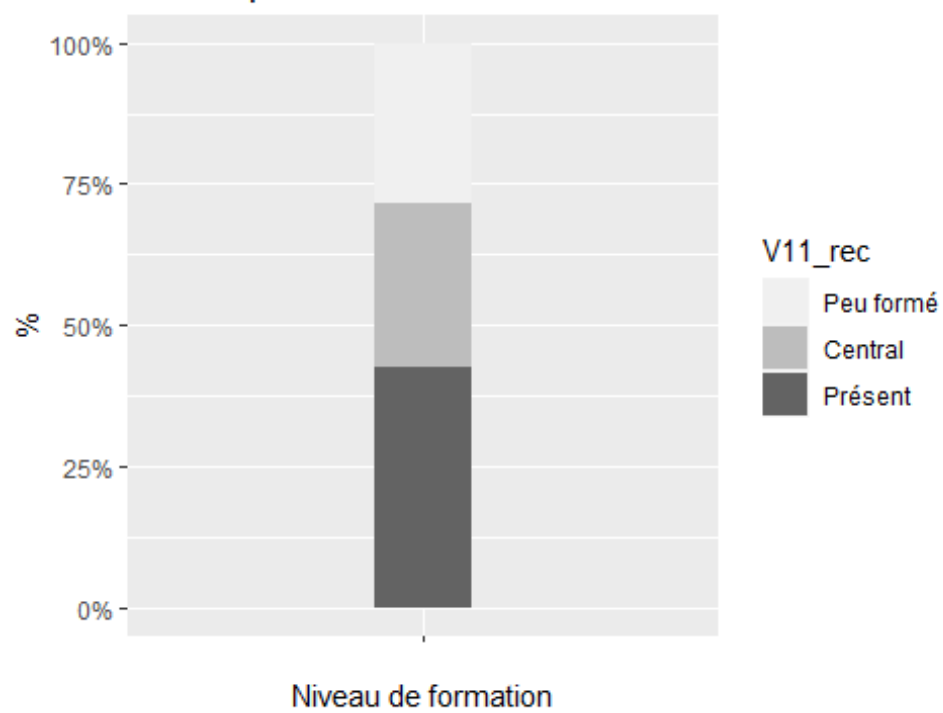
Manipulation de données au travail



La manipulation de données fait partie de l'activité professionnelle d'un peu plus de la moitié des répondants

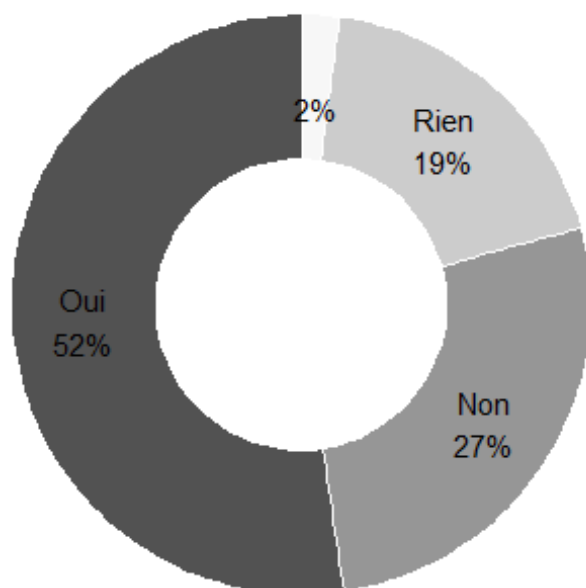
Mais parmi eux, 28% n'ont pas ou peu été formés à la manipulation de données

Ils manipulent des données au travail



Slide 6 : Les répondants arrivent en majorité à trouver ce qu'ils cherchent

Avez-vous trouvé ce que vous cherchiez ?



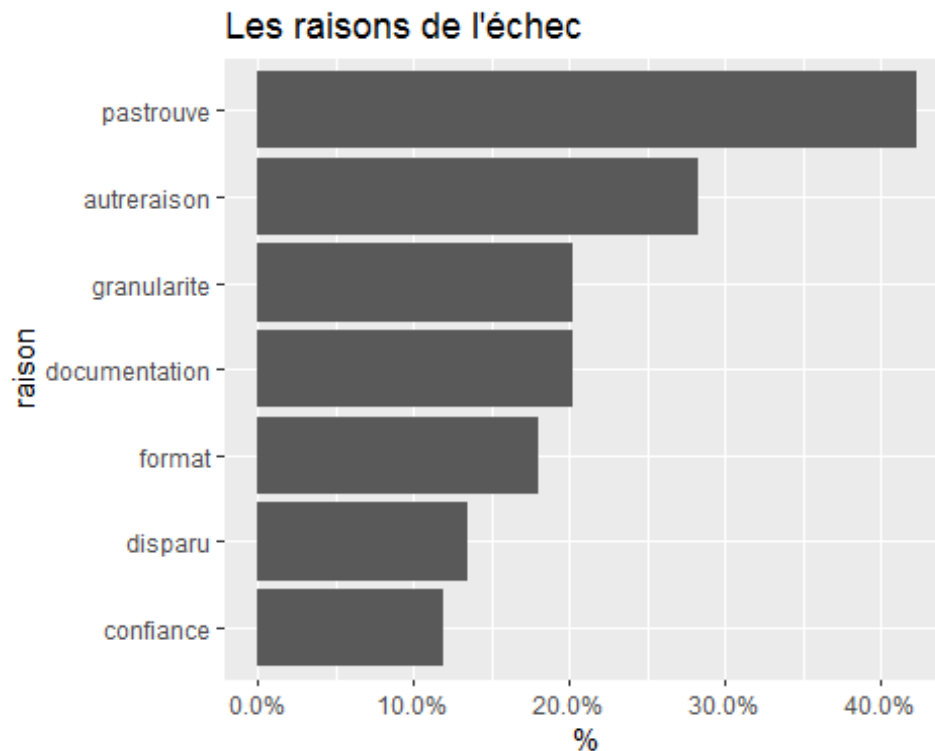
La moitié des répondants indiquent avoir trouvé ce qu'ils cherchaient sur data.gouv.fr

19 % ne cherchaient rien

27 % des répondants indiquent que leur recherche s'est soldée par un échec

NB : interprétation délicate des réponses à cette question, qui aurait demandé à être précisée. Par exemple : "avez-vous trouvé ce que vous cherchiez *lors de votre dernière visite* ?"

Slide 7 : Quand les recherches se soldent par un échec



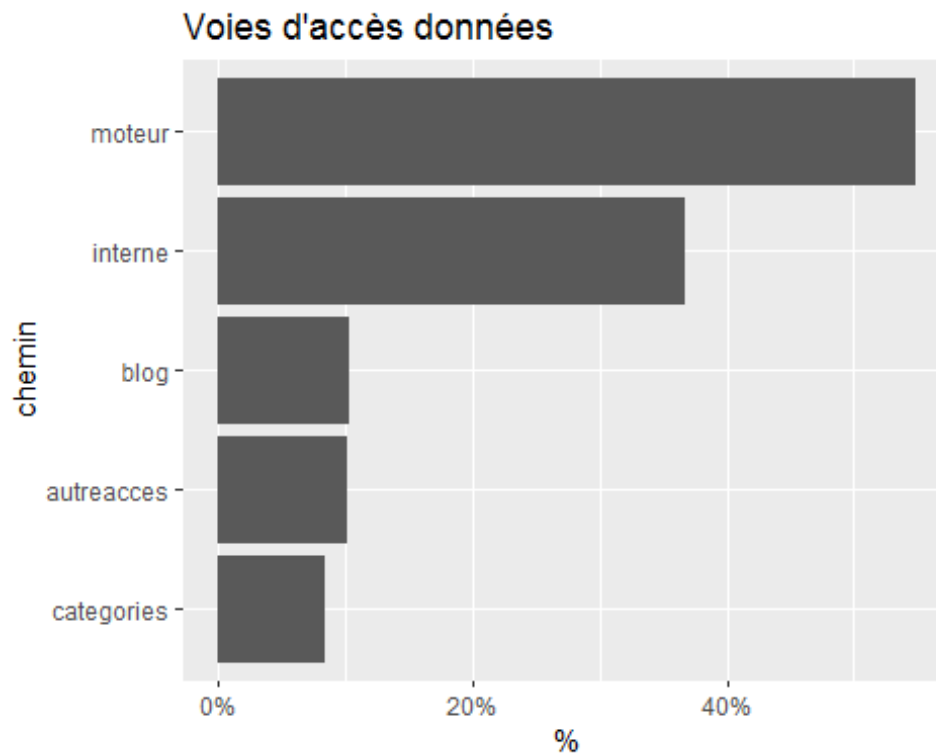
La raison la plus souvent avancée pour expliquer l'échec est : "le jeu de données n'existait pas ou je n'ai pas réussi à le retrouver via le moteur de recherche, les catégories"

Vient ensuite la catégorie "autre"

Ceux qui ont répondu "autre" précisent le plus souvent que c'est parce que les données étaient datées, n'étaient pas actualisées : "données cherchées trop anciennes", "données non actualisées", données obsolètes", "jeu de données pas à jour", "trop ancien"

NB : la question "Pourquoi ?" n'était pas filtrée, alors qu'on suppose qu'elle était implicitement dédiée aux individus qui avaient répondu "non" à la question précédente. Du fait de l'absence de filtre, certains des individus ayant répondu "oui" ou "je ne cherchais rien de particulier" ont également répondu à cette question. Au total, 371 individus ont répondu à la question (alors que seulement 243 avaient répondu "non" à la question précédente). On a décidé de garder toutes les réponses, considérant qu'elles doivent probablement être vues comme une liste de tous les problèmes déjà rencontrés par les utilisateurs de data.gouv.fr.

Slide 8 : Un accès principal aux données via un moteur de recherche externe



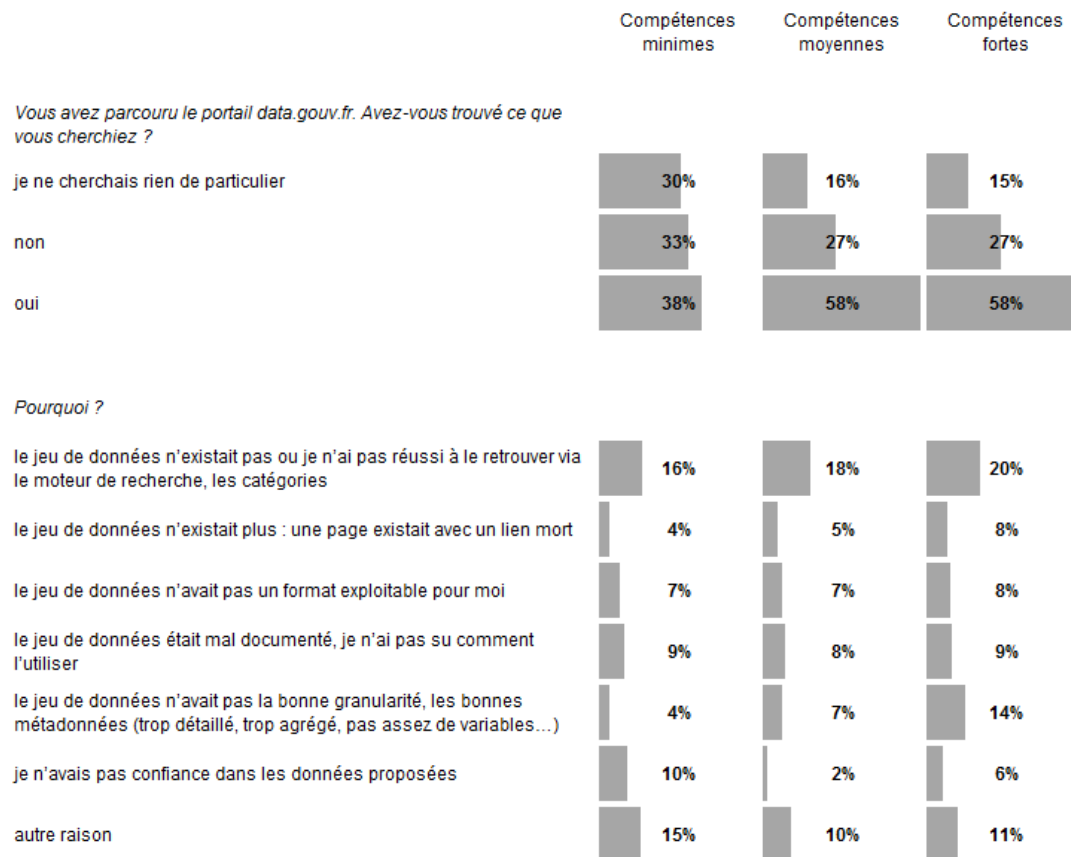
39% des répondants disent avoir trouvé le jeu de données via le moteur interne de data.gouv.fr

59% via un moteur de recherche externe

Ceux qui ont répondu "autre" précisent le plus souvent qu'ils sont arrivés au jeu de données via un autre site

NB : ici également la question aurait demandé à être précisée. On ne sait pas bien à quoi "ce jeu de données" fait référence, et à qui s'adressait cette question. Probablement à ceux ayant répondu "oui" à la question "Vous avez parcouru le portail data.gouv.fr. Avez-vous trouvé ce que vous cherchiez ?". Du fait de l'absence de filtre, une part importante des individus ayant répondu "non" ou "je ne cherchais rien de particulier" a également répondu. Au total, 840 individus ont répondu à la question. On a décidé de garder toutes les réponses, considérant qu'elles doivent probablement être vues comme les différentes voies d'accès aux données auxquelles ont recours les utilisateurs de data.gouv.fr.

Slide 9 : Qui trouve, qui ne trouve pas son jeu de données ?



Les répondants qui estiment avoir des compétences minimales en manipulation de données trouvent un peu moins ce qu'ils cherchent

Ils déclarent aussi plus souvent qu'ils ne recherchaient rien de particulier

Ils utilisent un peu moins le moteur de recherche interne de data.gouv.fr que les autres

En ce qui concerne les problèmes rencontrés, on ne note pas beaucoup de différences en fonction du niveau de compétences. Ceux qui ont des compétences minimales déclarent un peu plus fréquemment ne pas avoir confiance dans les données, et ceux qui ont des compétences fortes citent des problèmes liés à la granularité.

Slide 10 et suivantes : Retours d'utilisateurs sur la découvrabilité des données

Sélection des verbatims par Samuel