



HAL
open science

Recherches participatives, innovation ouverte et science ouverte Résultats de l'enquête nationale

Kenneth Maussang, H el ene Jouguet, Thomas Jouneau, Jean-Fran ois Martin,
Nicolas Larrousse

► To cite this version:

Kenneth Maussang, H el ene Jouguet, Thomas Jouneau, Jean-Fran ois Martin, Nicolas Larrousse. Recherches participatives, innovation ouverte et science ouverte R esultats de l'enqu ete nationale : R esultats de l'enqu ete nationale. Comit e pour la science ouverte. 2023. hal-04074860

HAL Id: hal-04074860

<https://hal-lara.archives-ouvertes.fr/hal-04074860>

Submitted on 19 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.



Distributed under a Creative Commons Attribution 4.0 International License



Recherches participatives, innovation ouverte et science ouverte

Résultats de l'enquête nationale

Kenneth Maussang
Hélène JOUGUET
Thomas JOUNEAU
Jean-François MARTIN
Nicolas LARROUSSE

Mars 2023

Recherches participatives, innovation ouverte et science ouverte

Résultats de l'enquête nationale

Comité pour la Science Ouverte

Collège Données de la recherche

MAUSSANG Kenneth
Université de Montpellier

Hélène JOUGUET
Huma-Num (UAR 3598), CNRS, Université d'Aix-Marseille, Campus Condorcet

Thomas JOUNEAU
Université de Lorraine

Jean-François MARTIN
Institut Agro, Montpellier SupAgro

Nicolas LARROUSSE
Huma-Num (UAR 3598), CNRS, Université d'Aix-Marseille, Campus Condorcet

Les rédacteurs de ce document s'expriment en leur nom propre et non au titre de leur employeur.

Mars 2023

- Les données de l'enquête sont disponibles sur l'entrepôt générique Recherche Data Gouv :

Maussang, Kenneth; Jouguet, Hélène; Jouneau, Thomas; Larrousse, Nicolas, 2023, "Enquête nationale "Recherches participatives, open innovation et science ouverte"", <https://doi.org/10.57745/Z4RPDK>, Recherche Data Gouv, V1, UNF:6:MTcVA/k71IszL1DqIhRtxQ== [fileUNF]

- Les codes sources utilisés pour analyser les résultats de l'enquête sont disponibles :

https://github.com/kmaussang/RP_enquete_analyse

Conception graphique : opixido



Except where otherwise noted, this work is licensed under <https://creativecommons.org/licenses/by/4.0/>

Résumé

Dans un projet de recherches participatives, des contributeurs non-scientifiques, dont le nombre peut être important, vont intervenir à une ou plusieurs étapes du processus de recherche. La communauté de participants ainsi formée est également hétérogène, tant du point de vue de la rigueur d'exécution des protocoles et des tâches qui leurs sont confiés, que dans leur investissement en temps et sur la durée du projet. La prise en compte de l'implication de ces non-scientifiques dans un tel projet nécessite donc de l'anticipation afin de s'assurer de la fiabilité des données produites ainsi que de la qualité des résultats scientifiques obtenus dans ce contexte.

Dans le cadre de ses travaux, le collège Données de la recherche du Comité pour la Science Ouverte a réalisé une enquête sur les pratiques des chercheurs et des services d'accompagnement, relatives à la production ou l'utilisation de données ouvertes dans le contexte des recherches participatives. L'accent a été mis sur la qualité des données et métadonnées obtenues dans cette démarche, relevant tout autant de la Science Ouverte que de l'Innovation ouverte. Cette enquête s'adressait à toute personne impliquée dans une activité de recherches participatives, quelques soient ses fonctions (chercheur, ingénieur, doctorant, technicien, service de soutien, membre d'une association,...), issue du secteur public, privé ou associatif (y compris les associations professionnelles, de malades ou de victimes).

Cette enquête a été menée en ligne du 28 novembre 2021 au 30 janvier 2022, et a rassemblé 359 réponses complètes, dont 24,5% des répondants ont déjà organisé un tel type de projet. Parmi ces derniers, 96,4% souhaitent renouveler l'expérience, alors que 44,1% des autres répondants envisagent d'organiser leur premier projet de recherches participatives. Ceci démontre un réel intérêt pour ce sujet, mais dont la mise en œuvre nécessite des dispositions/mesures particulières que l'enquête a permis d'identifier.

Principales conclusions

Parmi les répondants qui ont organisé un projet de recherches participatives,

- 30,7% déclarent avoir utilisé un plan de gestion de données ;
- 55,2% déclarent avoir eu une stratégie préalable pour assurer la qualité des données, 60,5% a posteriori, et 27,9% seulement pour assurer la qualité des métadonnées ;
- La formation des participants (non scientifiques) durant le projet est l'élément qui ressort comme étant le plus important et fréquemment cité ;
- Des scores de confiance dynamique permettent d'identifier les participants défaillants (recoupement avec d'autres participants ou un expert).

Bénéfices sur les données : les répondants ont déclaré que les démarches de recherches participatives apportent des bénéfices sur leurs données tels que :

- Augmentation de la quantité de données ;
- Plus large variété géographique de données ;

- Plus large variété sociologique ;
- Plus large variété des conditions expérimentales.

Certains répondants mentionnent que ces démarches participatives permettent de collecter des données difficilement accessibles autrement.

Ouverture des données, dissémination des résultats et communication : si 30,9% des répondants affirment ouvrir leurs données systématiquement ou souvent et 25,1% parfois, aucune différence notable n'a été constatée entre les répondants ayant déjà organisé un projet de recherches participatives et les autres au regard de l'ouverture de leurs données de recherche. Il a été constaté un faible recours aux licences pour la diffusion.

L'utilisation de licences associées aux jeux de données n'est pas systématique. Les précautions prises lors de l'ouverture de leurs données consistent principalement à les accompagner de notices explicatives de différentes natures : description des conditions d'acquisition, des protocoles, du contexte, des traitements appliqués aux données, et de la méthodologie. A noter que deux répondants ont mentionné l'utilisation de chartes d'utilisation des données.

Parmi l'ensemble des répondants, 63,5% déclarent avoir une stratégie de communication des résultats de leurs activités vers le grand public ou une démarche de médiation scientifique.

Méthodologie – profil des répondants

Les cibles de ce questionnaire étaient constituées de tous les acteurs du processus de recherche potentiellement impliqués dans un projet de recherches participatives, y compris le secteur associatif et le secteur privé (innovation ouverte, recherche partenariale). Le terme « qualité » n'a pas été défini volontairement afin de permettre l'appréciation individuelle au regard de la spécificité disciplinaire éventuelle. Seules les réponses complètes ont été analysées (359 réponses complètes pour 630 engagements). Les répondants sont issus très majoritairement d'établissements publics (92,5%), bien que quelques réponses d'associations (5,3%) et d'entreprises (2,2%) aient été collectées. La répartition disciplinaire déclarée permet d'obtenir des réponses issues de tous les champs disciplinaires, avec une prévalence des groupes CNU 10 (Biologie, biochimie, physiologie, neurosciences, écologie) et 9 (mécanique, génie informatique, automatique, traitement du signal, génie électrique, électronique, photoniques). Parmi les répondants, 88 ont déjà organisé un projet de recherches participatives et 71 d'entre eux ont indiqué un groupe CNU de rattachement (avec au moins une réponse par groupe CNU). Près de la moitié est issue des groupes 4 (psychologie, sociologie, histoire, géographie, urbanisme, ethnologie, philosophie, architecture) et 10.

Table des matières

Résumé	4
1 Contexte et méthodologie	7
1.1 Contexte	7
1.2 Cibles et objectifs	9
1.3 Profil des répondants	10
2 Rapport aux recherches participatives	14
3 Qualités des données et métadonnées issues d'un projet de recherche participatives	16
3.1 Plan de gestion de données	16
3.2 Stratégie préalable pour assurer la qualité des données	16
3.3 Stratégie a posteriori pour assurer la qualité des données	17
3.4 Précautions prises pour assurer la qualité des métadonnées	19
3.5 Bénéfices sur les données	19
4 Dissémination des résultats et communication	21
4.1 Stratégie de communication	21
4.2 Précautions lors de la communication des résultats	23
4.3 Lien avec les scolaires	25
5 Conclusions	26
Références	28
Liste des figures	28
Liste des tableaux	29

1 | Contexte et méthodologie

1.1 Contexte

Dans le cadre de ses travaux, le collège Données de la recherche du Comité pour la science ouverte a réalisé une enquête sur les pratiques des chercheurs et des services d'accompagnement, relatives à la production ou l'utilisation de données ouvertes dans le contexte des recherches participatives, de l'innovation ouverte et de la science ouverte, notamment la qualité des données et métadonnées dans ce type de démarche. Cette enquête s'adressait à toute personne impliquée dans une activité de recherche, d'innovation ou de recherches participatives, quel que soit ses fonctions (chercheur, ingénieur, doctorant, technicien, service de soutien, membre d'une association,...), issue du secteur public, privé ou associatif (y compris les associations professionnelles, de malades ou de victimes).

En préambule, des définitions des termes « recherches participatives », « innovation ouverte » et « R&D » étaient proposées comme suit :

Recherches participatives : les recherches participatives sont « des formes de production de connaissances scientifiques auxquelles des acteurs non scientifiques-professionnels — qu'il s'agisse d'individus ou de groupes — participent de façon active et délibérée »¹. Les groupes de personnes peuvent être des associations, ou des élèves d'écoles primaires, de collèges ou de lycées.

Innovation ouverte : l'innovation ouverte consiste à innover de manière ouverte, notamment grâce à des idées et des expertises venues de l'extérieur de l'entreprise².

R&D : recherche et développement.

Le terme « qualité » n'a pas été défini volontairement afin de laisser à l'appréciation du répondant sa propre appréciation des éléments qui déterminent la qualité d'un jeu de donnée, quel que soit les champs disciplinaires. Le questionnaire était disponible du 30 novembre 2021 au 31 janvier 2022, accessible via un lien internet, implanté avec la solution logiciel LimeSurvey. Initialement la date limite de réponse était au 12 janvier 2022, reporté ensuite au 31 janvier 2022. Les répondants avaient la possibilité de sauvegarder leurs réponses pour reprendre le remplissage plus tard. Sur cette période de deux mois, 630 personnes ont initié une réponse sur le questionnaire, dont 359

¹ François Houllier, Sciences Participatives en France. Etats des lieux, bonnes pratiques et recommandations, février 2016, <https://hal.inrae.fr/hal-02801940>

² <https://www.bpifrance.fr/nos-actualites/quatre-facons-de-faire-de-lopen-innovation>

réponses complètes et validées. A noter que 34 répondants de plus ont complété le formulaire jusqu'à la visualisation de la dernière page du questionnaire sans le valider. Seules les réponses validées et envoyées ont été analysées dans ce document. La durée moyenne de réponse des répondants est de 7min et 10s quand le temps de réponse indicatif était annoncé à 15min maximales. 94% des répondants ont répondu en moins de 15min, 87% en moins de 10min et 46% en moins de 5min. Certaines sections du questionnaire étaient conditionnées aux réponses précédentes, expliquant cette disparité de temps de réponse.

Les répondants avaient la possibilité de ne pas rendre publique leurs réponses et/ou leur établissement d'appartenance s'ils le souhaitaient. Au final, 149 répondants ne souhaitent pas rendre public le nom de leur établissement (42,45%). A titre de comparaison, une enquête en ligne avait été menée dans le cadre de la rédaction du rapport Houllier sur les sciences participatives³. Cette dernière avait obtenu 608 réponses, dont 188 complètes, avec un temps moyen de réponse supérieur à une heure. Parmi ces 188 réponses complètes, 63% étaient fournies par des scientifiques (153 réponses).

Cette enquête a été complétée d'un court questionnaire diffusé auprès du réseau SupDPO afin d'obtenir un retour d'expérience des délégués à la protection des données (DPO) d'établissements d'enseignement supérieur sur ces projets. Il visait notamment à identifier d'éventuelles stratégies de respect de la RGPD. Ce questionnaire spécifique n'a toutefois pas trouvé son public, avec deux réponses seulement. Si l'on met de côté la stratégie perfectible de diffusion de ce questionnaire, peut-être peut-on voir en creux dans ce taux de réponse très bas, le faible niveau d'identification du lien entre les projets de recherches participatives et les questions de RGPD. A noter que l'auteur de l'une des deux réponses souhaite inclure à l'avenir les questions de recherches participatives dans les formations aux doctorants sur les données personnelles qu'il dispense.

Suite à une erreur de paramétrage dans le schéma conditionnel du questionnaire, deux questions sur l'utilisation de licences dans la diffusion de données ouvertes n'ont pas été affichées aux participants. Les données de cette enquête sont disponibles⁴.

³ François Houllier, Sciences Participatives en France. Etats des lieux, bonnes pratiques et recommandations, février 2016, <https://hal.inrae.fr/hal-02801940>

⁴ Maussang, Kenneth; Jouguet, Hélène; Jouneau, Thomas; Larrousse, Nicolas, 2023, "Enquête nationale "Recherches participatives, open innovation et science ouverte"", <https://doi.org/10.57745/Z4RPDK>, Recherche Data Gov, V1, UNF:6:MTcVA/k71IszL1DqIhRtxQ== [fileUNF]

Figure 1- Nombre de réponses quotidiennes complètes et cumul durant la période où le questionnaire était accessible en ligne (du 30/11/2021 au 31/01/2022).

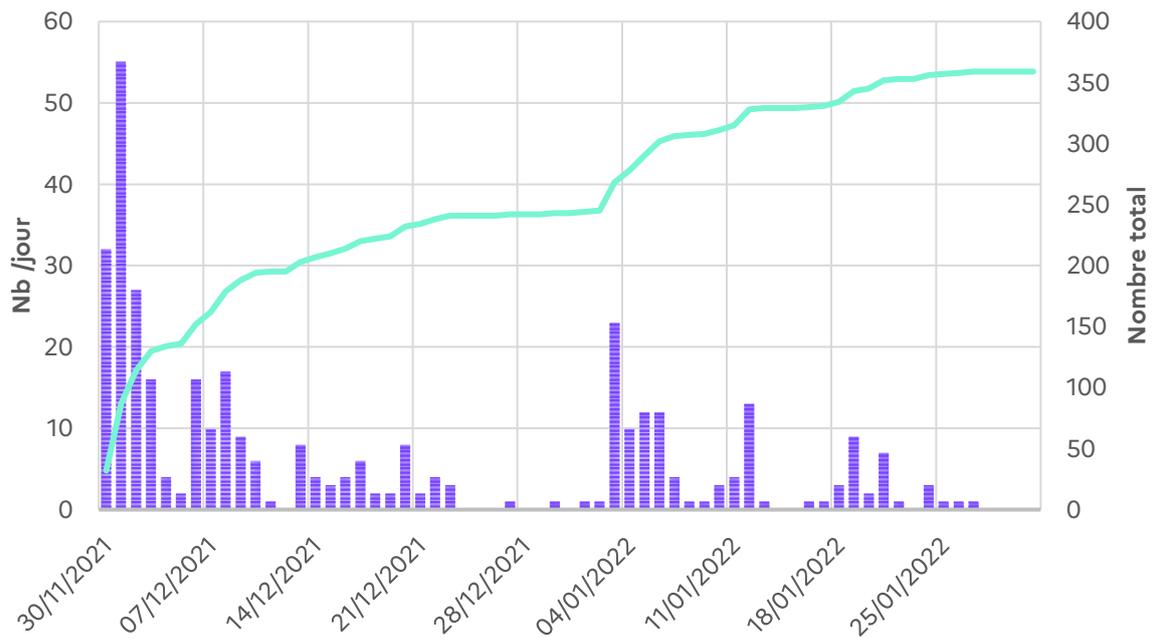
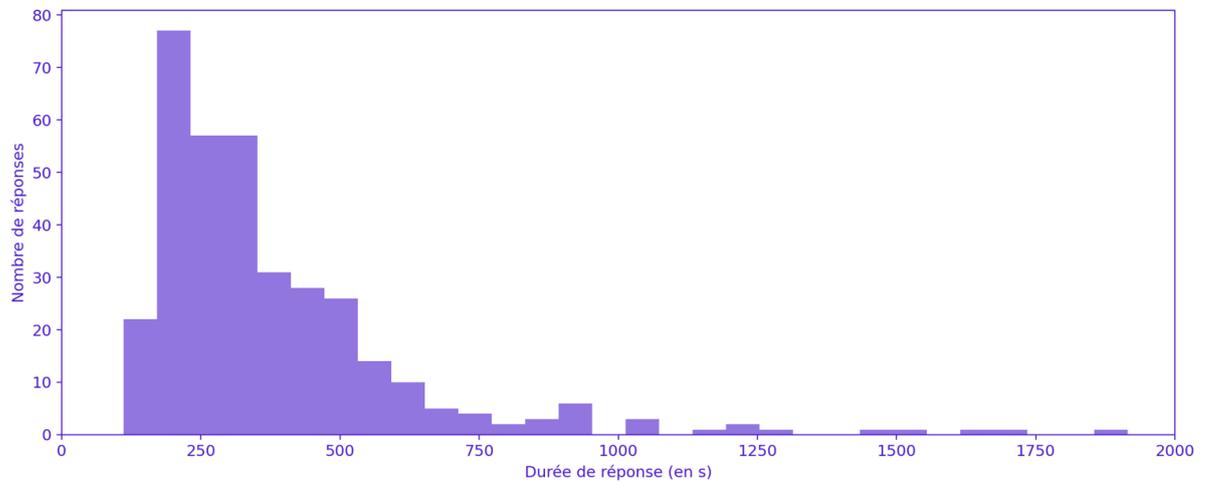


Figure 2- Histogramme des temps de réponse des répondants.



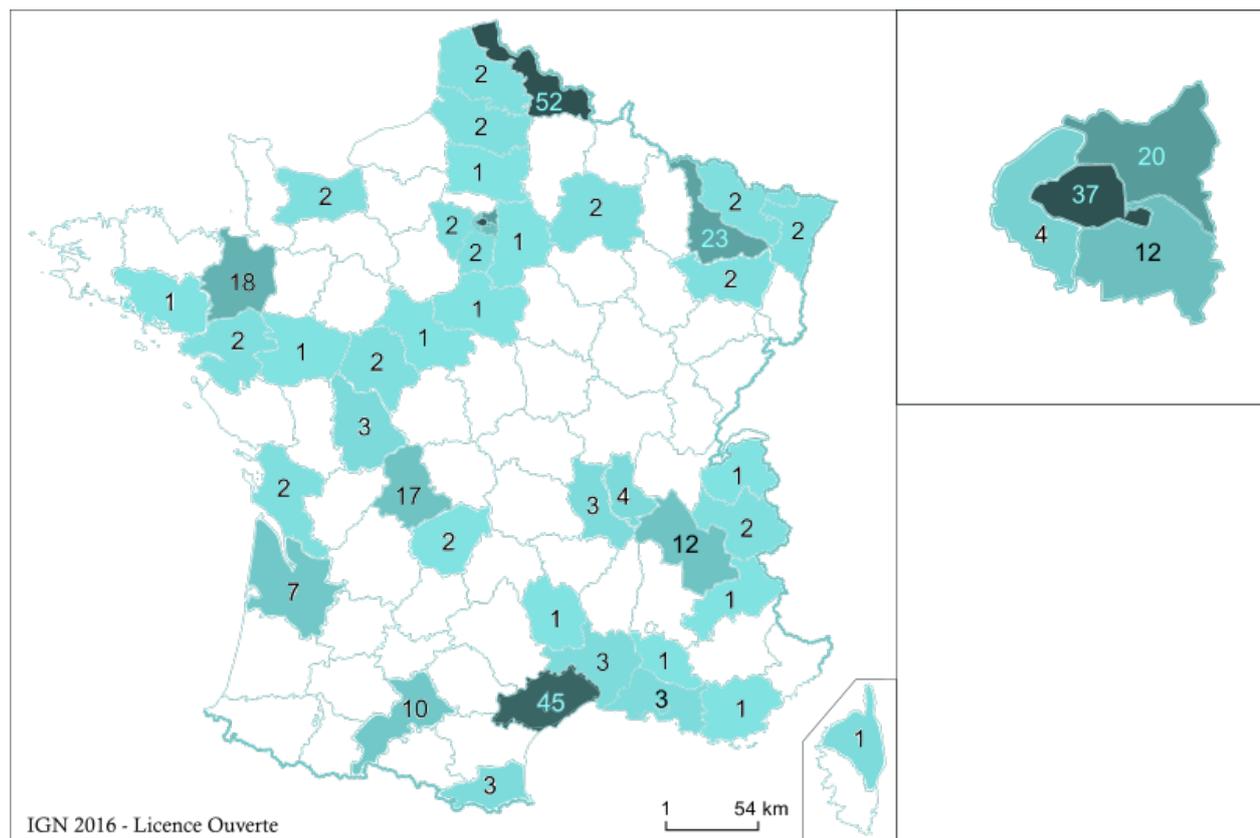
1.2 Cibles et objectifs

L'objectif de cette enquête consiste à obtenir un retour d'expérience des acteurs des projets de recherches participatives sur les difficultés éventuellement rencontrées pour assurer la production ou l'analyse de données de qualité dans ce type de projets, les protocoles qualité mis en œuvre le cas échéant, ou les précautions méthodologiques mises en œuvre.

Les cibles de ce questionnaire étaient constituées de tous les acteurs du processus de recherche potentiellement impliqués dans ce type de projet, et également les services de soutien à la recherche, y compris le secteur associatif et le secteur privé engagé dans des démarches d'innovation ouverte ou de recherche partenariale.

1.3 Profil des répondants

Figure 3 - Répartition géographique des répondants en France métropolitaine. Fond de carte issu de l'IGN sous Licence Ouverte (<https://www.data.gouv.fr/fr/datasets/fonds-de-carte-ign-france-et-regions-571459/>).



Si tous les champs disciplinaires sont représentés, cette enquête n'a cependant aucunement la prétention de fournir une image représentative de tous les points de vue des acteurs concernés et de tous les champs disciplinaires ou de significativité statistique. Les résultats produits permettent de dégager une première tendance, aucune étude similaire n'ayant par ailleurs été conduite. Les répondants sont issus très majoritairement d'établissements publics (92,48%), bien que quelques réponses d'associations (5,29%) et d'entreprises (2,23%) aient été collectées. Parmi les répondants issus d'établissements publics, 62,35% sont issus d'établissements d'enseignement supérieur public (université, grande école, ...), 31,02% d'organismes de recherches publics de type EPST (CNRS, INRAE, INRIA, INSERM,...), 2,11% d'EPIC et 4,52% d'autres établissements publics. Sur l'ensemble des répondants, 42,45% ne souhaitent pas rendre public le nom de leur établissement, mais certains d'entre eux ont accepté qu'il soit anonymisé, ce qui permet de connaître la nature des établissements des répondants. Il est ainsi possible de noter que 82 établissements ne sont mentionnés qu'une seule fois (23,36% des réponses). Les fonctions exercées par les répondants sont variées, avec

majoritairement des enseignants-chercheurs permanents (32,6%), des chercheurs permanents (19,5%) et des doctorants (14,5%). Parmi les répondants se déclarant affiliés à une association (19 répondants), seuls 3 appartiennent à une association de professionnels. Près des trois quarts des répondants ont un statut de permanent ou assimilé dans leurs fonctions (60,5% de fonctionnaires et 13,17% de CDI).

Tableau 1- Fonctions exercées par les répondants.

Fonction	Nombre de répondants	
Enseignant-chercheur permanent	117	32.59%
Chercheur permanent	70	19.50%
Doctorant	52	14.48%
Ingénieur d'étude	25	6.96%
Autre	21	5.85%
Ingénieur de recherche	20	5.57%
Chef de projet, chargé de projet	16	4.46%
Post-doctorant	14	3.90%
Technicien	7	1.95%
Responsable R&D	4	1.11%
Chargé de mission SAPS (Science Avec et Pour la Société)	4	1.11%
Informaticien	4	1.11%
Bibliothécaire	4	1.11%
Chargé d'affaire ou équivalent dans un service de valorisation	1	0.28%

Avec 23,25% de répondants en CDD, les bénévoles ou répondants sans contrat restent marginaux. En termes de niveau d'étude, 72,35% des répondants sont titulaires d'un niveau BAC+8 (doctorat, HDR) et 23,74% d'un niveau BAC+5 (master, diplôme d'ingénieur, ...). Les disciplines déclarées révèlent une répartition des réponses issues de tous les champs disciplinaires, avec une prévalence des domaines de l'écologie, la biologie mais également l'électronique ou le traitement du signal correspondant⁵ aux groupes disciplinaires CNU-10 et CNU9 (intitulés CNU – Conseil National des Universités - décrit dans le tableau ci-dessous). Les projets de recherches participatives ont précédemment été étudiés⁶ et une prévalence des projets en écologie notamment avait été constatée.

Les recherches participatives permettent d'accroître la quantité de données d'observation, d'assurer une grande variété géographique et une collecte dans le temps long, et sont donc adaptées aux recherches en environnement, tout particulièrement en écologie. Si l'on restreint aux répondants déclarant avoir déjà porté un projet de recherches participatives, l'ensemble des groupes disciplinaires restent représentés,

⁵ Le champ disciplinaire était à renseigner à partir de la liste des groupes CNU défini au sens du conseil national des universités (<https://conseil-national-des-universites.fr/cnu>), sans préjuger de l'établissement d'appartenance du répondant qui ne se restreignait pas aux établissements d'enseignement supérieur et de recherche.

⁶ European Commission. Joint Research Centre., Survey report: data management in Citizen Science projects. LU: Publications Office, 2016. Consulté le: 9 février 2022. [En ligne]. Disponible sur: <https://data.europa.eu/doi/10.2788/539115>

même sur les groupes 1, 6, 8 et 12 du CNU sont faiblement représentés (1 ou 2 répondants).

Tableau 2- Définition des groupes disciplinaires CNU.

Section	Intitulé
CNU-1	Droit privé, droit public, sciences politiques, histoire du droit (sections 1 à 4)
CNU-2	Sciences économiques, sciences de gestion et du management (sections 5 et 6)
CNU-3	Littérature, langues, sciences du langage, linguistique (sections 7 à 15)
CNU-4	Psychologie, sociologie, histoire, géographie, urbanisme, ethnologie, philosophie, architecture (sections 16 à 24)
CNU-5	Mathématiques, informatique (sections 25 à 27)
CNU-6	Physique hors astrophysique/astronomie (section 28 à 30)
CNU-7	Chimie (sections 31 à 33)
CNU-8	Astronomie/astrophysique, géologie, météorologie (sections 34 à 37)
CNU-9	Mécanique, génie informatique, automatique, traitement du signal, génie électrique, électronique, photoniques (sections 60 à 63)
CNU-10	Biologie, biochimie, physiologie, neurosciences, écologie (sections 64 à 69)
CNU-12	Pluridisciplinaire (sections 70 à 74)

Tableau 3- Répartition disciplinaire des répondants.

Groupe CNU	Nombre de répondants	
CNU-10	77	23.19%
CNU-9	36	10.84%
CNU-4	33	9.94%
CNU-7	31	9.34%
Santé	28	8.43%
CNU-5	25	7.53%
CNU-6	19	5.72%
CNU-3	15	4.52%
CNU-2	15	4.52%
Service de soutien aux chercheurs	14	4.22%
CNU-12	12	3.61%
CNU-8	8	2.41%
Interdisciplinaire	7	2.11%
Autre	6	1.81%
CNU-1	5	1.51%
Théologie	1	0.30%

Si 88 répondants ont déjà organisé un projet de recherches participatives, 71 d'entre eux ont indiqué un groupe CNU de rattachement. Les non-répondants correspondent soit à des membres d'associations ou entreprises privées, soit une absence de réponse à la question correspondante. Parmi les répondants ayant déjà organisé un projet de recherche participative, près de la moitié sont issus des groupes CNU 4 et 10 (respectivement 25,35% et 22,54%). Cela n'est guère étonnant pour le groupe CNU 10 compte-tenu de l'importance déjà constatée des projets de recherches participatives

en écologie/environnement⁷. Le groupe 4 du CNU comprend les recherches en sociologie qui expliquent l'importance des parts de répondants appartenant à ce groupe-là.

Tableau 4- Répartition disciplinaire des répondants ayant déjà organisé un projet de recherches participatives.

Groupe CNU	Nombre de répondants (ayant déjà porté un projet de recherches participatives)	
CNU-4	18	25.35%
CNU-10	16	22.54%
Santé	6	8.45%
CNU-3	5	7.04%
CNU-2	5	7.04%
CNU-9	4	5.63%
CNU-5	4	5.63%
CNU-7	3	4.23%
Interdisciplinaire	2	2.82%
Autre	2	2.82%
CNU-6	2	2.82%
CNU-12	2	2.82%
CNU-8	1	1.41%
CNU-1	1	1.41%

Cependant, il convient de souligner qu'un projet de recherche participative implique une participation active d'un non-scientifique et que ces derniers ne sont pas « simplement » un objet de mesure ou de collecte de données tels que des capteurs ou de systèmes de collecte de données passives qu'un individu aurait consenti d'utiliser dans son environnement usuel. Dans un projet de recherches participatives, les participants non scientifiques au projet sont des parties prenantes actives du processus de recherche.

La variété géographique des répondants peut être appréciée au regard de la ville déclarée, synthétisée Figure 4.

⁷ European Commission. Joint Research Centre., Survey report: data management in Citizen Science projects. LU: Publications Office, 2016. Consulté le: 9 février 2022. [En ligne]. Disponible sur: <https://data.europa.eu/doi/10.2788/539115>

Tableau 5 - Niveau de maîtrise des projets de recherches participatives.

Avez-vous déjà organisé un projet de recherches participatives ?	Comment évaluez-vous votre niveau de maîtrise des projets de recherches participatives ?	Nombre de répondants	
Non	Aucune connaissance sur le sujet	78	28.8%
	Notions	158	58.3%
	Maîtrise	31	11.4%
	Expert(e)	4	1.5%
Oui	Aucune connaissance sur le sujet	0	0.0%
	Notions	16	18.2%
	Maîtrise	50	56.8%
	Expert(e)	22	25.0%

Tableau 6 - Intention d'organiser un projet de recherches participatives dans un futur proche.

Avez-vous déjà organisé un projet de recherches participatives ?	Songez-vous, dans un futur proche (quelques années), organiser un projet de recherches participatives ? Si oui, dans quel délai ?	Nombre de répondants	
Non	Non	146	55.9%
	Oui, dans les semaines/mois qui viennent	12	4.6%
	Oui, dans 1 à 2 ans	30	11.5%
	Oui, dans plus de 2 ans	47	18.0%
	Autre	26	10.0%
Oui	Non	3	3.6%
	Oui, dans les semaines/mois qui viennent	47	56.0%
	Oui, dans 1 à 2 ans	23	27.4%
	Oui, dans plus de 2 ans	5	6.0%
	Autre	6	7.1%

3 | Qualités des données et métadonnées issues d'un projet de recherche participatives

3.1 Plan de gestion de données

Parmi les répondants qui ont organisé un projet de recherches participatives, seulement 30,68% déclarent avoir utilisé un plan de gestion de données. Si 57,95% déclarent ne pas en avoir utilisé, 11,36% ne savent pas et donc il est vraisemblable qu'ils n'en aient pas utilisé.

3.2 Stratégie préalable pour assurer la qualité des données

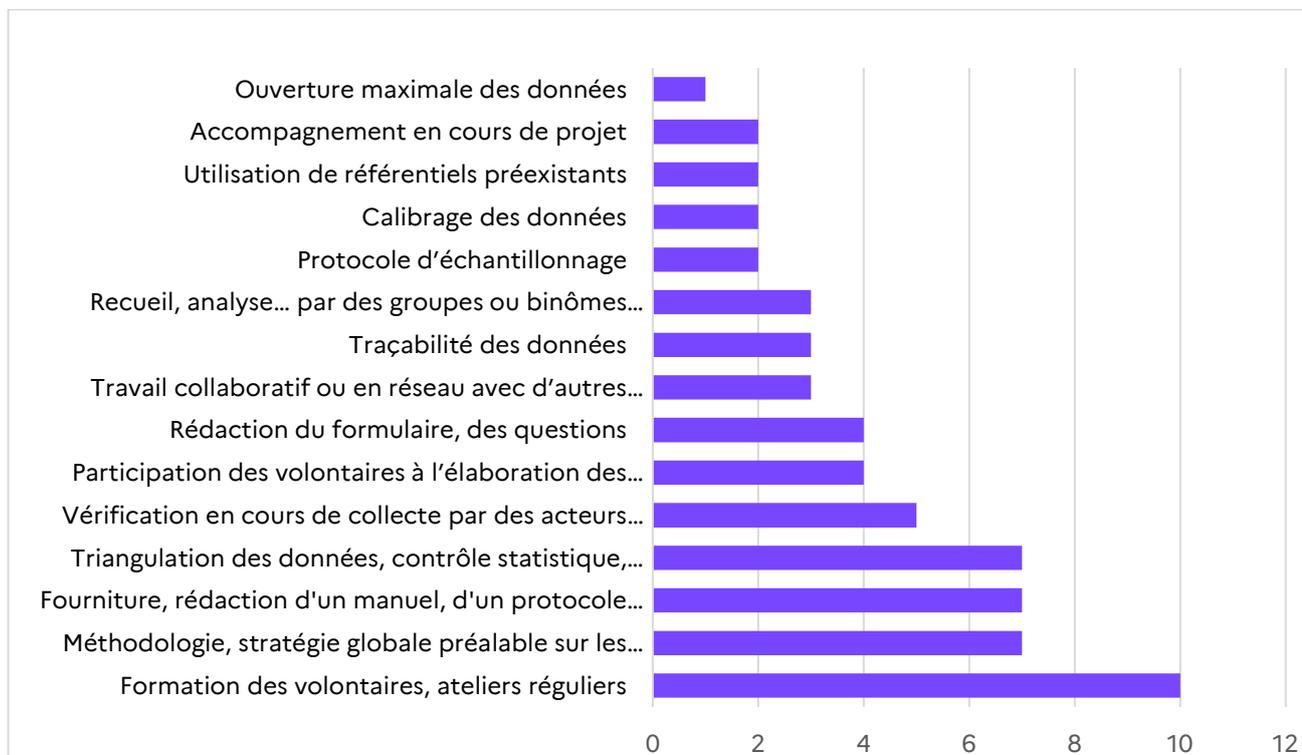
Parmi les répondants qui ont organisé un projet de recherches participatives, 55,17% déclarent avoir eu une stratégie pour assurer la qualité des données. Pour assurer la qualité des données, les participants ont principalement proposé les stratégies préalables décrites Figure 5.

Les points suivants ont par ailleurs été mentionnés :

- Collaboration avec d'autres acteurs scientifiques ;
- Enregistrement et retranscription systématiques des entretiens en profondeur et des réunions menées avec les acteurs ;
- Kit d'échantillonnage ;
- Design méthodologique ;
- Stockage sur Huma-Num ;
- Choix de moyens techniques sécurisés ;
- Test de plusieurs protocoles ;
- Mise en place d'un comité de pilotage ;
- Stockage provisoire et transfert des prélèvements ;
- Prétests et pas d'analyse des données avant la fin de toutes les passations ;
- Echantillon d'experts de communauté ;

- Base de données pour la traçabilité des prélèvements effectués par les participants.

Figure 3 - Principales stratégies mise en place pour assurer la qualité des données a priori. Agglomération de réponses proposées en champ de texte libre.



3.3 Stratégie a posteriori pour assurer la qualité des données

Parmi les répondants qui ont organisé un projet de recherches participatives, 60,47% déclarent avoir eu une stratégie pour assurer la qualité des données *a posteriori*. La liste des réponses fournies en champ de texte libre est disponible section Pour assurer la qualité des données dans un projet de recherches participatives, les répondants ont principalement proposés les démarches suivantes :

- Score de confiance dynamique attribué aux participants ;
- Comparaison des résultats entre experts et participants sur un sous-ensemble ;
- Relecture par les experts et feedback aux participants (en lien avec la formation) ;
- Définition des critères de qualité ;
- Identification des données suspectes et validation par un expert le cas échéant ;
- Identification des participants systématiquement défaillant (produisant des résultats suspects).

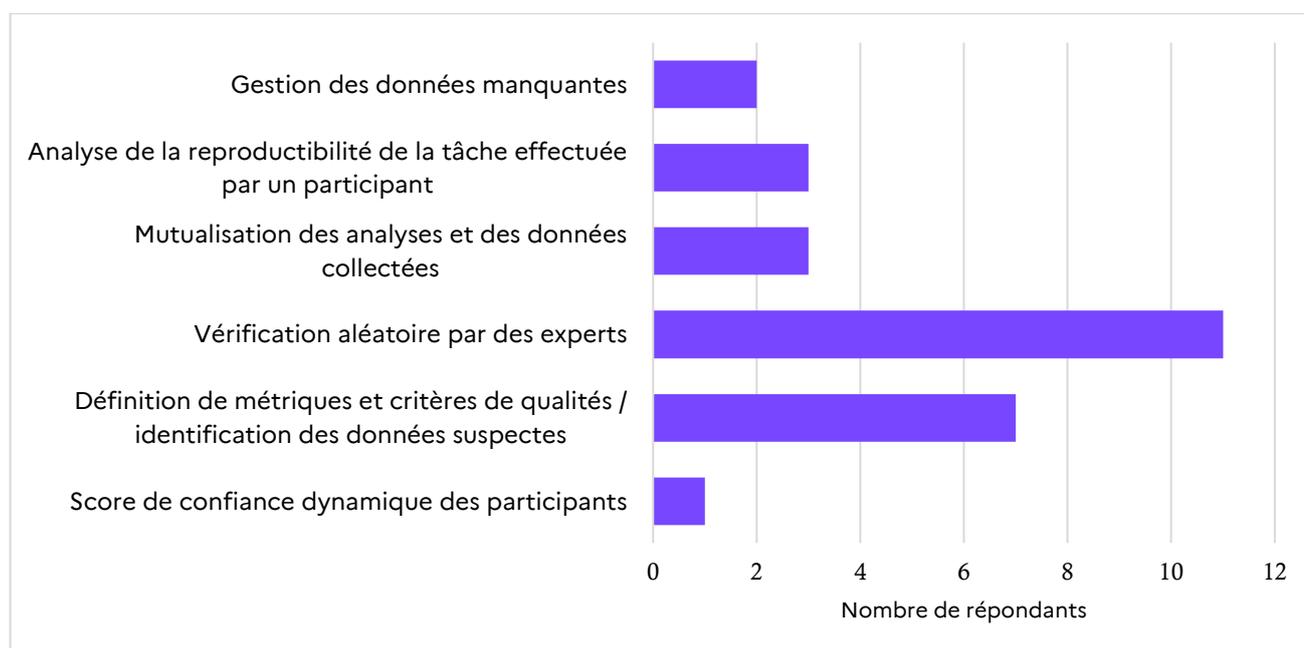
Dans le cas spécifique d'une analyse de données par les participants :

- Proposer le même jeu de données à plusieurs participants pour comparer les résultats de l'analyse ;
- Proposer un jeu de données connu/calibré (analysé par un expert) pour « tester » la qualité du participant ;
- Vérification aléatoire ;
- Comparaison des résultats entre experts et participants sur un sous-ensemble ;
- Dans le cas des transcriptions, comparaison des résultats sur des segments communs du corpus ;
- Analyse de la reproductibilité de l'analyse des données : des données similaires analysées par le même participant.

Dans le cas spécifique d'une collecte de données par les participants :

- Regroupement des données produites par les participants (valeurs moyennes, etc...) ;
- Définition de métriques de qualité dans le cas d'une collecte à l'aide de capteurs (rapport signal sur bruit par exemple) ;
- Analyse de la reproductibilité des données : des données similaires produites par plusieurs participants dans des conditions similaires ;
- Définition des indicateurs de cohérence si cela est pertinent ;
- Gestion des données manquantes.

Figure 4 - Principales stratégies mise en place pour assurer la qualité des données a posteriori. Agglomération de réponses proposées en champ de texte libre.



Nous pouvons catégoriser les principales mesures prises comme suit :

- Mise en œuvre de comparaisons et vérifications en utilisant des résultats d'enquêtes similaires
- Recours à l'expertise de spécialistes pour contrôler les résultats
- Contrôle de cohérence automatisé par la plate-forme ou par recoupements « logiques » suivant les disciplines.

Les procédures de contrôle varient manifestement selon la discipline concernée, le recours à la comparaison et analyse d'écart est par exemple manifestement largement répandu et habituel pour certains répondants.

3.4 Précautions prises pour assurer la qualité des métadonnées

Parmi les répondants qui ont organisé un projet de recherches participatives, seulement 27,91% déclarent avoir pris des précautions pour assurer la qualité des métadonnées, soit une proportion similaire à la section précédente. Les répondants précisent avoir pris les mesures suivantes :

- Rédaction de guides avec pour certains un accompagnement des participants ;
- Définition de protocoles de collecte ;
- Recours à des outils qui permettent la structuration des données et l'usage de vocabulaires contrôlés ;
- Vérification *a posteriori* (par des experts).

Il semble que pour certains répondants les trois questions se recoupent en partie, aussi certaines réponses se recoupent avec les questions précédentes. Cela semble être le cas là où la nature du projet permet un contrôle réparti tout au long du projet sans forcément distinguer les étapes comme précisé dans les questions.

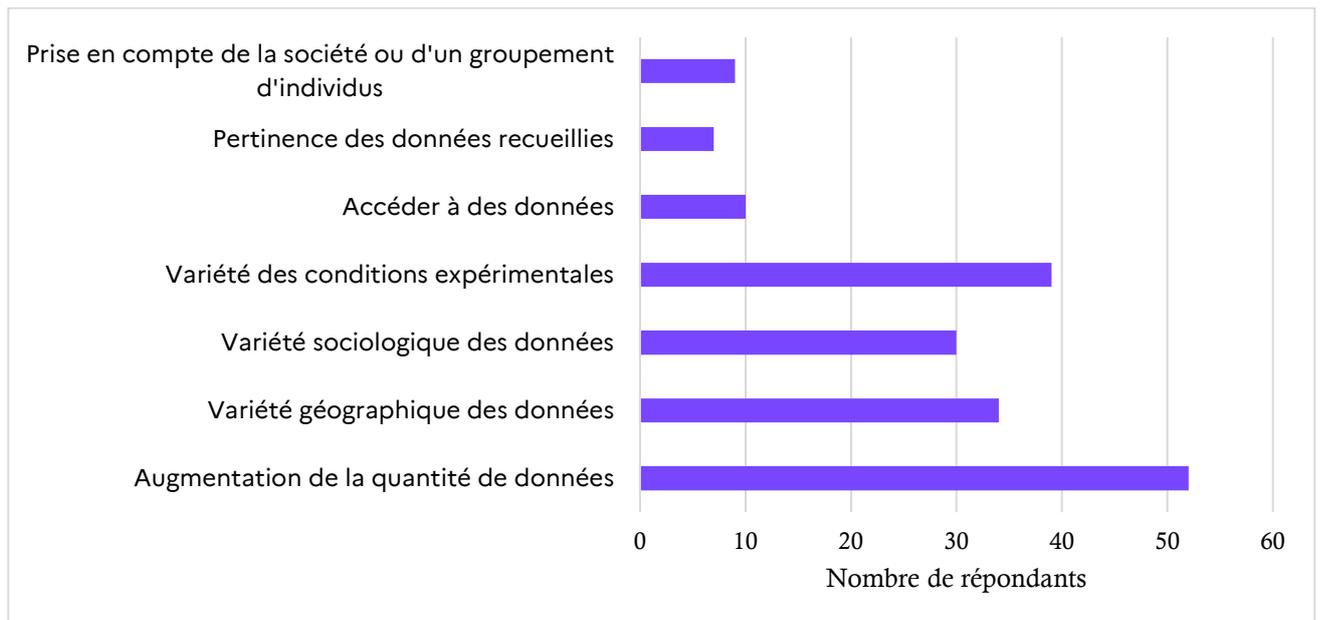
3.5 Bénéfices sur les données

Les répondants ont confirmé que les démarches de recherches participatives permettent notamment d'accroître la quantité de données produites, leur variété géographique ou sociologique par exemple, mais également une variété des conditions expérimentales.

Tableau 7 - Bénéfices d'une démarche de recherches participatives sur les données produites ou analysées. La liste des réponses « autre » fournies en champ de texte libre est disponible section 8, page 19.

Bénéfice sur les données d'une démarche de recherches participatives	Nombre de répondants	
Augmentation de la quantité de données	52	74.3%
Variété géographique des données	34	48.6%
Variété sociologique des données	30	42.9%
Variété des conditions expérimentales	29	41.4%
Autre	36	51.4%

Figure 5 - Principales bénéfices d'une démarche participative sur les données.
Agglomération de réponses proposées en champ de texte libre.



Plusieurs répondants ont mis en avant un intérêt pour accéder à des données difficiles d'accès ou augmenter la quantité des données. D'autres répondants mettent en avant un intérêt sur la pertinence des données recueillies et/ou leur qualité. Certains répondants mettent en avant la pertinence d'une démarche participative pour la prise en compte de la société civile ou de groupement d'individus

- « Le premier bénéfice est la prise en compte de questions de recherche inédites, venant de la société civile, l'interdisciplinarité, les comparaisons internationales »
- « Réponse à une question de recherche posée par une association de malades »
- « Croiser les savoirs des professionnels, des chercheurs et des personnes concernées, trianguler les données (sources, méthodes, analyse) »
- « Multiples bénéfices sans aucun lien avec les données : utilité et pertinence de la recherche, connaissance des besoins du monde professionnel... »

4 | Dissémination des résultats et communication

L'ouverture des données est généralement conseillée car elle peut être stimulante pour les participants, dès lors que les conditions légales et éthiques sont respectées⁸. Elle favorise la transparence du dispositif et de son suivi par les acteurs qui souhaitent s'impliquer. Dans un projet de recherches participatives, la donnée est un élément de communication important. Il ne faut pas voir la donnée comme une matière première pour la phase d'analyse mais d'en exploiter pleinement son potentiel de communication pour la dissémination des connaissances⁹. Par conséquent, l'enquête s'est intéressée aux pratiques en termes d'ouverture de données et de dissémination des résultats des répondants.

4.1 Stratégie de communication

Parmi l'ensemble des répondants, 63,53% déclarent avoir une stratégie de communication des résultats de leurs activités vers le grand public ou une démarche de vulgarisation. Les canaux de communications utilisés sont variés, mais restent principalement l'utilisation d'un site internet institutionnel (68,1% des répondants), Twitter (33,4% des répondants) et LinkedIn (29,3% des répondants), bien avant l'utilisation de communiqués de presse (23,7%). Les web TV restent d'utilisation marginale (2,2%) en regard des médias locaux (22,1%), nationaux (18%) ou les journaux scientifiques grand public (24,6%). L'utilisation de Chaînes Youtube (ou équivalent) sont en revanche bien plus répandus (16,7%).

Tableau 8 - Cibles de la stratégie de communication des répondants.

Cibles de la stratégie de communication	Nombre de répondants	
Utilisateur final / clients ou pool de clients	3	1.4%
Collègues ou confrères	185	84.1%
Actionnaires	1	0.5%
Grand public	179	81.4%
Journalistes	77	35.0%
Elus	46	20.9%
Autre	44	20.0%

⁸ F. Houllier et J.-B. Merilhou-Goudard, « Les sciences participatives en France », Rapport élaboré à la demande des ministres en charge de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche, 2016. doi: 10.15454/1.4606201248693647E12.

⁹ L. C. S. W. Group, Citizen Science Skilling for Library Staff, Researchers, and the Public. LIBER Citizen Science Working Group, 2021. doi: 10.25815/hf0m-2a57

Tableau 9 - Canaux de communication utilisés.

Canaux de communication utilisés	Nombre de répondants	
Site internet institutionnel	216	68.1%
Twitter	106	33.4%
LinkedIn	93	29.3%
Facebook	60	18.9%
Tiktok	2	0.6%
Instagram	18	5.7%
Média(s) locaux	70	22.1%
Média(s) nationaux	57	18.0%
Web tv	7	2.2%
Chaîne Youtube ou équivalent	53	16.7%
Journal scientifique grand public (epsilon,...)	78	24.6%
Communiqué de presse	75	23.7%
Autre	94	29.7%

Parmi les répondants, 43,1% affirment avoir déjà communiqué des éléments techniques à un journaliste scientifique pour la rédaction d'un article ou d'un reportage. Concernant les données produites dans le cadre de leurs activités, 30,9% des répondants affirment publier leurs données en open data systématiquement ou souvent, 25,1% parfois. En revanche, 34% affirment ne pas le faire et 10% ne savent pas. Aucune différence notable n'a été constatée entre les répondants ayant déjà organisé un projet de recherches participatives et les autres. Les répondants disposaient d'un champ de texte libre pour indiquer les précautions qu'ils prenaient pour accompagner la diffusion publique de leurs données.

Tableau 10 - Ouverture des données.

Avez-vous déjà organisé un projet de recherches participatives ?	Publiez-vous les données issues de vos activités en open data ?	Nombre de répondants	
Non	Non	87	33.0%
	Oui, parfois	70	26.5%
	Oui, souvent	47	17.8%
	Oui, systématiquement	33	12.5%
	Je ne sais pas	27	10.2%
Oui	Non	32	37.2%
	Oui, parfois	18	20.9%
	Oui, souvent	16	18.6%
	Oui, systématiquement	12	14.0%
	Je ne sais pas	8	9.3%

4.2 Précautions lors de la communication des résultats

Dans une première question, les répondants ont indiqué sous forme de réponse en texte libre les précautions qu'ils prennent pour communiquer leurs résultats de manière générale, synthétisé Figure 8.

Un nombre important déclare ne prendre aucune précaution (25 répondants) alors que la principale précaution prise est l'accompagnement par une note explicative du contexte, du protocole, etc... (26 répondants). Le respect des règles éthiques est souvent mentionné (19 répondants), ainsi que des messages d'avertissement (11 répondants) et des notes légales (8 répondants). Certains répondants ont mis en avant la mention de l'établissement d'appartenance et/ou le financeur comme éléments de communication associé (7 répondants). Quelques répondants affirment ne communiquer qu'après avoir publié les résultats dans une revue. Enfin, deux répondants ont indiqué s'appuyer sur le service communication de leur établissement. Ensuite, les répondants pouvaient indiquer les précautions particulières prises dans le cas spécifique de la communication de leurs données (Figure 9). Ces précautions consistent principalement à les accompagner de notices explicatives de différentes natures. Les différents types de notices explicatives proposés sont ventilés Figure 10. L'essentiel de ces notices explicatives consiste à décrire les conditions d'acquisition, les protocoles, les traitements appliqués aux données, et la méthodologie. Des notes d'explication du contexte sont également proposées. Seulement 18 répondants ont mentionné l'utilisation de licences associées aux jeux de données, alors que 11 répondants suivent les recommandations de l'éditeur quand ils publient ou les recommandations de l'entrepôt où ils déposent les données. A noter que deux répondants ont mentionné l'utilisation de chartes d'utilisation des données.

Figure 6 - Précautions prises par les répondants pour diffuser leurs résultats.
Agglomération de réponses proposées en champ de texte libre.

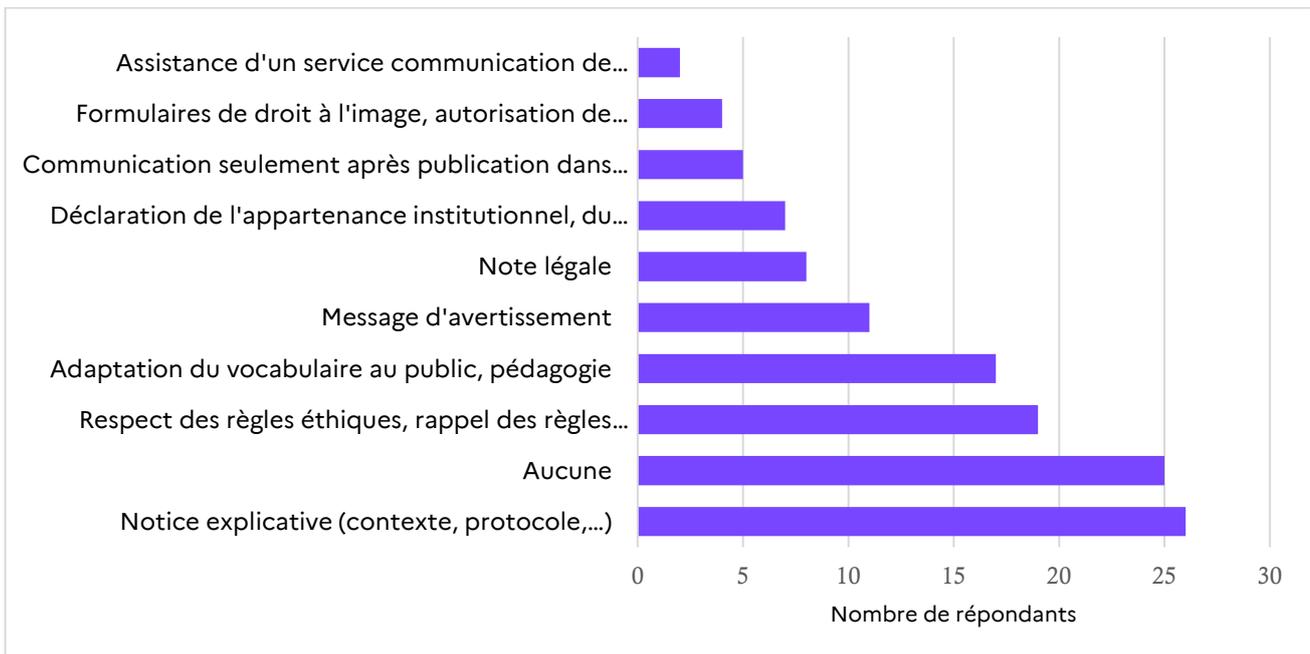


Figure 7 - Principaux types de précautions prises par les répondants pour diffuser leurs données. Les différents types de notices explicatives sont décrites plus en détails

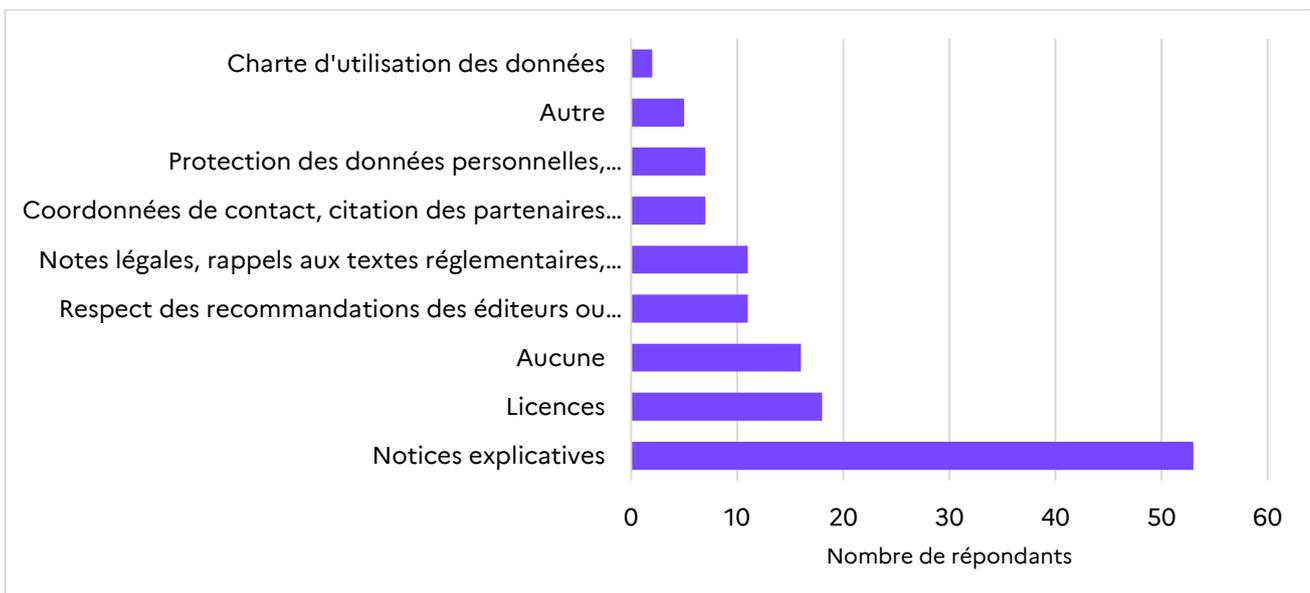
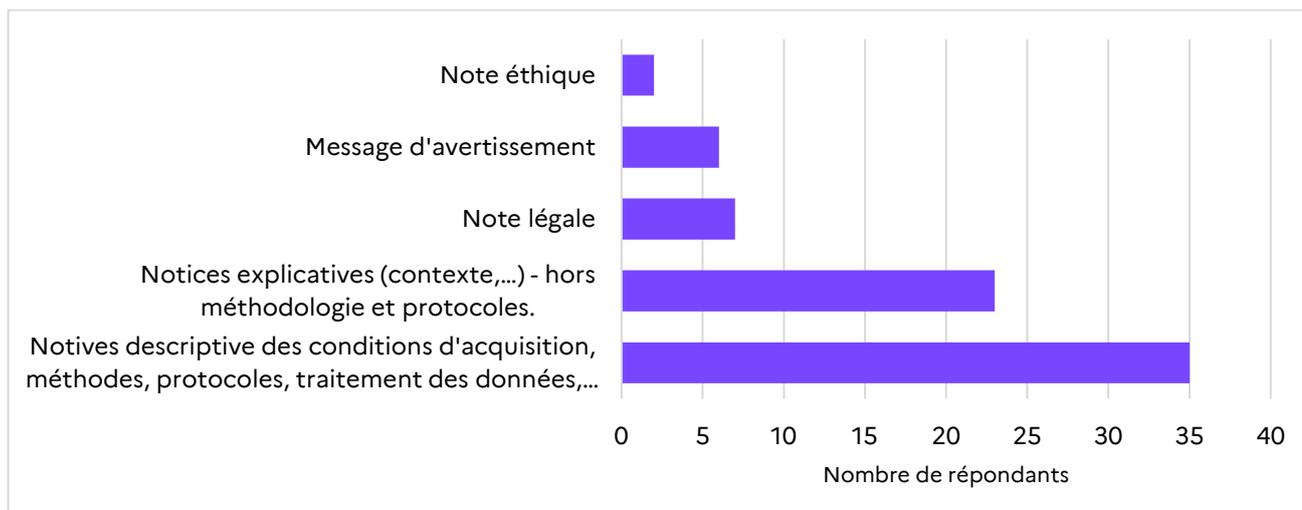


Figure 8 - Principaux types de notices explicatives proposées par les répondants pour diffuser leurs données. Agglomération de réponses proposées en champ de texte libre.



4.3 Lien avec les scolaires

Une part importante des répondants déclare travailler personnellement avec des scolaires dans le cadre de leurs fonctions (46,07%), y compris dans le cadre de la production de ressources pédagogiques. Les entreprises et associations déclarent également produire des ressources pédagogiques à destination des établissements supérieures, majoritairement pour les universités (56,5%) et les grandes écoles (26,1%). Dans ce cas, les ressources produites sont variées : données brutes (41,7%), données traitées (58,3%), données techniques (25%), dossiers pédagogiques (58,3%), code informatique (25%), vidéos (41,7%), images (66,7%), ... Ces résultats sont à nuancer compte tenu du faible nombre de répondants issus d'entreprises ou d'associations.

Tableau 11 - Ressources fournies aux établissements d'enseignement supérieurs par les associations ou entreprises des répondants.

Ressources	Nombre de répondants	
Données brutes	5	41.7%
Données traitées	7	58.3%
Données techniques	3	25.0%
Dossiers pédagogiques	7	58.3%
Code	3	25.0%
Vidéos	5	41.7%
Images	8	66.7%
Matériel de TP	2	16.7%
Autre	3	25.0%

5 | Conclusions

Avec 359 réponses complètes en seulement 2 mois, cette enquête démontre une nouvelle fois le dynamisme des démarches de recherches participatives¹⁰, même si seulement 24,5% des répondants ont déjà organisé un tel type de projet. Parmi ces derniers, 96,4% souhaitent renouveler l'expérience, alors que 44,1% des autres répondants envisagent d'organiser leur premier projet de recherches participatives. Ceci démontre un réel intérêt pour ce sujet, mais dont la mise en œuvre nécessite la prise en compte de l'intervention de non-scientifiques dans le processus de recherche. Les résultats de l'enquête font ressortir une série d'éléments clés à considérer dans cette perspective.

La communication en est un premier, qui permet de maintenir dans le temps l'engagement des participants, qui sont en attente d'un retour de la part des équipes de recherche impliquées. L'ouverture des données produites est alors un levier important de maintien de l'engagement et de l'intérêt des participants. Cependant aucune différence notable n'a été constatée entre les répondants ayant déjà organisé un projet de recherches participatives et les autres au regard de l'ouverture de leurs données de recherche. L'ouverture des données et la communication des résultats nécessitent de prendre les mêmes précautions que dans le cadre des démarches de science ouverte dans le cas général, notamment par l'utilisation de licences sur les jeux de données. Si l'utilisation de licences par les répondants de cette enquête a été mentionnée, les précautions prises lors de l'ouverture de leurs données consistent principalement à les accompagner de notices explicatives de différentes natures : description des conditions d'acquisition, des protocoles, du contexte, des traitements appliqués aux données, et de la méthodologie. A noter que deux répondants ont mentionné l'utilisation de chartes d'utilisation des données. En cas de collecte participative de données, l'utilisation d'une plateforme de dépôt de données associée par exemple à une cartographie en temps réelle et ouverte – sans temporisation – participe à la motivation et au sentiment de confiance des participants. Cette enquête n'a pas différencié l'ouverture des données et le partage restreint aux participants du projet. Au-delà de ce travail, il serait intéressant de s'intéresser aux spécificités des recherches participatives et des enjeux associés au partage des données restreint à la communauté de participants.

Parmi l'ensemble des répondants, 63,5% déclarent avoir une stratégie de communication des résultats de leurs activités vers le grand public ou une démarche de vulgarisation. Ces projets de recherches participatives ont par essence une composante d'ouverture à partir du moment où les processus de recherche sont ouverts à des non-scientifiques. La motivation des participants est maintenue par une contrepartie : montée en compétence via la formation, intérêt pour le sujet et donc intérêt pour les données produites lors du projet (impliquant une ouverture des données) ou communication des résultats à la communauté de participants.

¹⁰ Au sens défini par le présent document : dans un projet de recherches participatives, des participants non scientifiques d'un projet de recherche peuvent intervenir à n'importe quelle étape du processus de recherche scientifique : collecte de données, analyse de données, mais également lors de la construction de la problématique de recherche initiale.

Parmi les répondants qui ont déjà organisé un projet de recherches participatives, environ 31% déclarent avoir utilisé un plan de gestion de données (PGD). Ces projets ont été réalisés avant l'obligation réglementaire de production d'un PGD par les agences de financement (ANR, ERC,...) marquant ainsi la pertinence de l'outil pour de tels projets.

L'implication de non-scientifiques nécessite cependant deux précautions supplémentaires par rapport à un projet standard de recherche : proposer des notices explicatives (protocoles, etc...) accessibles au niveau de connaissance des participants, et proposer en documentation des jeux de données une notice explicative du contexte de production des données correspondant, mentionnant explicitement la contribution des participants non-scientifiques et selon quelles modalités. Ces précisions participent à la fiabilisation des données produites et permettent d'avoir les éléments de contexte nécessaires à leur exploitation correcte suivant les types de projets en cas de réutilisation par la suite (échantillonnage biaisé en fonction de la population de participants par exemple).

Différentes typologies de projets de recherches participatives existent avec une variété des contributions des participants. Ils peuvent contribuer à n'importe quelle étape du processus de recherche. Suivant le niveau d'intervention de ces derniers, les stratégies mises en place pour assurer la qualité et/ou la fiabilité des données produites ne seront pas nécessairement les mêmes. La majorité des répondants affirment avoir anticipé a priori et/ou a posteriori une stratégie préalable pour assurer la qualité des données. Anticiper les contraintes associées à l'intervention de participants non-scientifiques, en gros nombre, est nécessaire pour assurer une fiabilité suffisante des données produites et/ou des analyses sous-jacentes, et garantir la rigueur scientifique du processus de recherche. Effectuer des statistiques sur des grands nombres n'est pas toujours pertinent, si la qualité initiale est insuffisante.

La formation des participants (non-scientifiques) durant le projet reste l'élément qui ressort comme étant le plus important et fréquemment cité. La forme que prend cette formation peut être très variable suivant le type de projet cependant. L'autre stratégie consiste à développer un protocole suffisamment balisé et fermé pour que le participant n'ait que peu de marge de manœuvre et peu de risque de mal effectuer la tâche qui lui incombe. Un certain nombre de répondants mettent en avant aussi l'utilisation de scores de confiance dynamiques, permettant d'identifier les participants défaillants ou pondérer les différentes contributions. Dans le cas d'analyse de données participatives, il est également fréquemment cité la possibilité de faire analyser le même jeu de données par plusieurs participants et croiser les résultats, puis par échantillonnage aléatoire de croiser ces analyses avec celles d'un scientifique expert. D'autres approches consistent à proposer au non-scientifique un jeu de données connu ou calibré par un scientifique pour évaluer individuellement la qualité de la participation. Dans le cas de transcriptions de manuscrits, une méthode consiste à comparer les résultats sur des segments communs du corpus. Au final, les contributions à cette enquête mettent en avant l'intérêt de définir au préalable des métriques de qualité/fiabilité, soit des participants, soit de leur production dans le projet, quelle que soit la nature de leurs tâches. Les répondants ont également mis en avant le bénéfice sur leurs données des démarches de recherches participatives, notamment par l'augmentation de la quantité de données, de la variété géographique, de la variété sociologique ainsi que des conditions expérimentales. Certains répondants affirment également que ces démarches participatives sont bénéfiques parce qu'elles permettent d'accéder à des données difficilement accessibles autrement.

Références

F. Houllier et J.-B. Merilhou-Goudard, « Les sciences participatives en France », Rapport élaboré à la demande des ministres en charge de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche, 2016. doi: 10.15454/1.4606201248693647E12.

European Commission. Joint Research Centre., Survey report: data management in Citizen Science projects. LU: Publications Office, 2016. Consulté le: 9 février 2022. [En ligne]. Disponible sur: <https://data.europa.eu/doi/10.2788/539115>

L. C. S. W. Group, Citizen Science Skilling for Library Staff, Researchers, and the Public. LIBER Citizen Science Working Group, 2021. doi: 10.25815/hf0m-2a57.

M. D. Wilkinson et al., « The FAIR Guiding Principles for scientific data management and stewardship », *Sci. Data*, vol. 3, no 1, p. 160018, mars 2016, doi: 10.1038/sdata.2016.18.

Liste des figures

Figure 1- Nombre de réponses quotidiennes complètes et cumul durant la période où le questionnaire était accessible en ligne (du 30/11/2021 au 31/01/2022).	9
Figure 2- Histogramme des temps de réponse des répondants.	9
Figure 5 - Principales stratégies mise en place pour assurer la qualité des données a priori. Agglomération de réponses proposées en champ de texte libre.	17
Figure 6 - Principales stratégies mise en place pour assurer la qualité des données a posteriori. Agglomération de réponses proposées en champ de texte libre.....	18
Figure 7 - Principales bénéfices d'une démarche participative sur les données. Agglomération de réponses proposées en champ de texte libre.	20
Figure 8 - Précautions prises par les répondants pour diffuser leurs résultats. Agglomération de réponses proposées en champ de texte libre.	24
Figure 9 - Principaux types de précautions prises par les répondants pour diffuser leurs données. Les différents types de notices explicatives sont décrites plus en détails	24
Figure 10 - Principaux types de notices explicatives proposées par les répondants pour diffuser leurs données. Agglomération de réponses proposées en champ de texte libre.	25

Liste des tableaux

Tableau 1- Fonctions exercées par les répondants.....	11
Tableau 2- Définition des groupes disciplinaires CNU.....	12
Tableau 3- Répartition disciplinaire des répondants.	12
Tableau 4- Répartition disciplinaire des répondants ayant déjà organisé un projet de recherches participatives.....	13
Tableau 5 - Niveau de maîtrise des projets de recherches participatives.	15
Tableau 6 - Intention d'organiser un projet de recherches participatives dans un futur proche.....	15
Tableau 7 - Bénéfices d'une démarche de recherches participatives sur les données produites ou analysées. La liste des réponses « autre » fournies en champ de texte libre est disponible section 8, page 12.	19
Tableau 8 - Cibles de la stratégie de communication des répondants.	21
Tableau 9 - Canaux de communication utilisés.....	22
Tableau 10 - Ouverture des données.	22
Tableau 11 - Ressources fournies aux établissements d'enseignement supérieurs par les associations ou entreprises des répondants.	25