

# Débat Ouverture des données de santé *Open Data*

*Résultats de la Consultation publique*

**30 avril 2014**

# Les principales conclusions de la consultations

- La consultation publique a intéressé essentiellement des personnes initiées à l'open data et à la santé : **203 participants ont répondu précisément en sélectionnant des jeux de données spécifiques**
- Parmi les participants qui se sont identifiés, les répondants majoritaires sont des **cabinets de conseils, des associations, des assurances et des centres hospitaliers**
- Les 10 catégories des jeux de données les plus demandées sont les suivantes :
  - **Epidémiologie**
  - **Etat de santé**
  - **Inégalités de santé**
  - **Tarifs et honoraires**
  - **Infrastructures**
  - **Services proposés**
  - **Activités des établissements**
  - **Consultations**
  - **Médicament et dispositifs**
  - **Dépenses**
- Les 10 jeux de données les plus demandés sont :
  - **Des données anonymisées du PMSI**
  - **Des données anonymisées du DCIR**
  - **La base des causes de mortalité**
  - **Les tableaux des affections longue durée**
  - **Les tableaux des « honoraires des professionnels de santé »**
  - **Les fichiers « analyse des dépenses »**
  - **Données sur l'offre de soin des cartographies du SNATIH**
  - **Des données anonymisées sur DMP**
  - **Des données anonymisées de l'EGB**
  - **Le suivi des honoraires des médecins par spécialité et lieu d'Exécution**
- Pour les participants, les objectifs principaux pour l'ouverture des données de santé sont « **la transparence, la confiance et la démocratie** » puis « **ressource pour l'innovation économique et sociale** ». De nombreux commentaires soulignent l'importance de l'ouverture des données publiques en santé pour **renforcer la recherche clinique, améliorer la veille sanitaire ainsi que l'innovation**
- **Les résultats de la consultation sont publiés sur le site data.gouv.fr**

# Les chiffres de la consultation

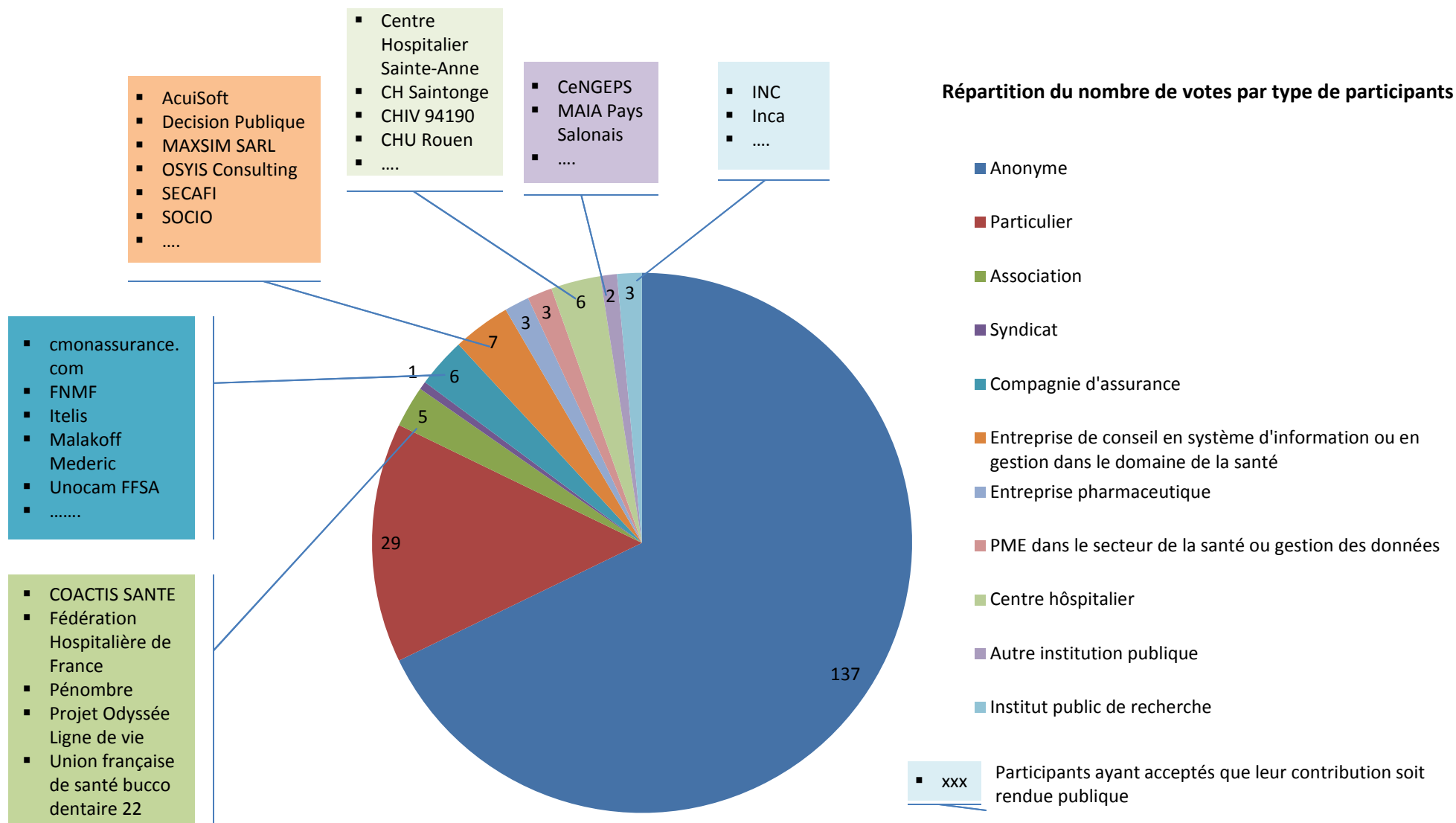
## Modalités

- Questionnaire en ligne accessible à partir des sites du SGMAP et de la Commission Open Data (site de la DREES)
- Une communication faite sur le site d' Etalab et dans certains médias (émission des « Clics et des claques » sur Europe 1 du 4 avril, article du monde le 8 avril)
- Réponses recueillies durant 3 semaines en avril 2014
- Possibilité donnée aux utilisateurs de s'identifier et de rendre publique leur contribution
- 4 grandes catégories de données
  - Des sous-ensembles par catégories
  - Des jeux de données ligne à ligne
  - Certains jeux se rapportant à plusieurs catégories proposés plusieurs fois (ex: DCIR du SNIIRAM, PMSI)

## Résultats

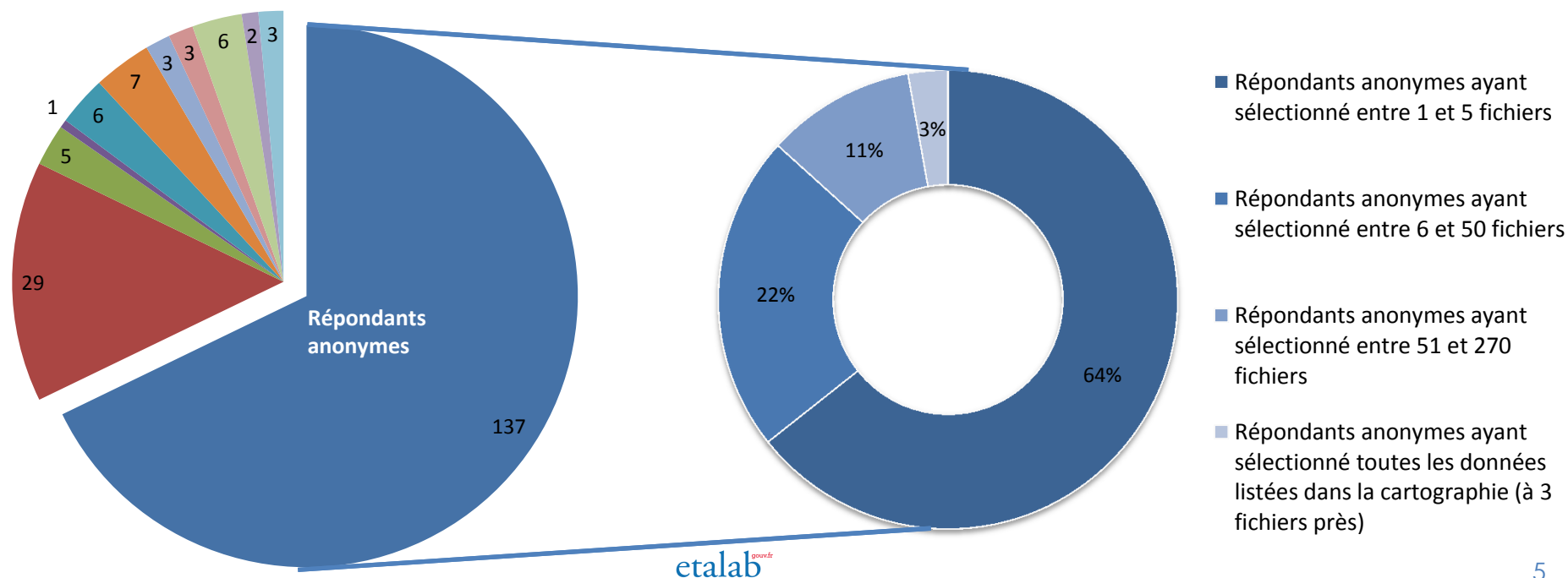
- Questionnaire ouvert **1 700** fois
- **326** questionnaires ont été remplis :
  - dont **203** sont considérés comme « achevés »
  - et **123** pour pour lesquels seules les catégories de données ont été renseignées (questionnaires inachevés ou choix délibéré)
- Au total, **plus de 7 000 « votes » exprimés** pour l'ouverture de données spécifiques

# Sur les 203 questionnaires finalisés, 66 participants se sont identifiés et 137 ont souhaité rester anonymes



# Les répondants anonymes ont majoritairement voté pour 1 à 2 données à ouvrir

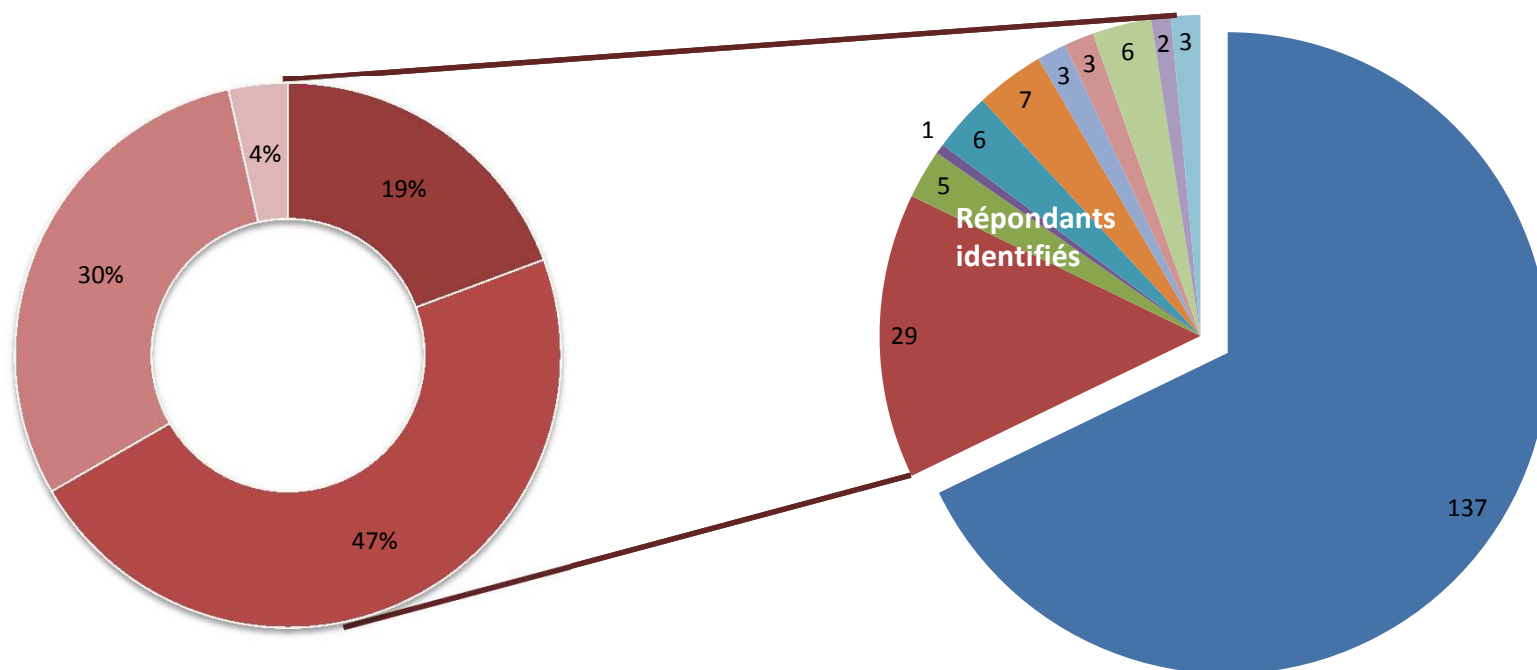
- Les anonymes sont probablement bien informés sur les données de santé :
  - La majorité des répondants anonymes ont fait une sélection ciblée sur 1 à 2 fichiers dont ils demandent l'ouverture
  - Ces anonymes ont rarement sélectionné les mêmes fichiers (récurrence maximum de 4 sélections pour un fichier donné parmi ces répondants)
- Il y a proportionnellement autant de « convaincus » de l'open data parmi les anonymes et que les répondants identifiés :
  - Quatre répondants anonymes ont sélectionné un à un tous les fichiers de la cartographie



# Les répondants identifiés ont sélectionné en moyenne un plus grand nombre de données à ouvrir

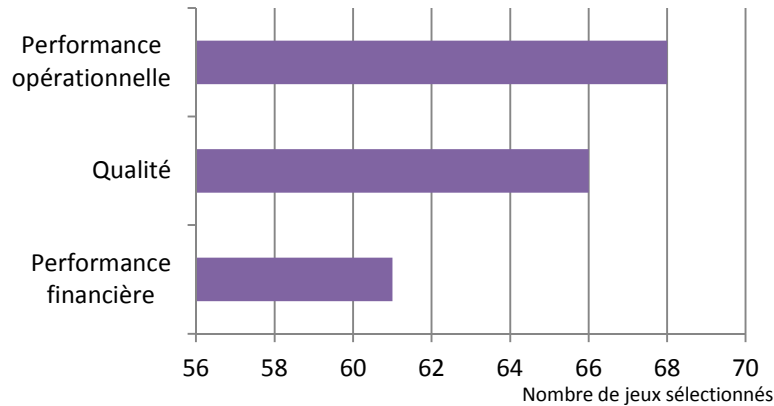
- Les répondants qui se sont identifiés (particuliers et institutionnels) ont sélectionné 54 fichiers à ouvrir en moyenne (médiane = 19 fichiers), soit deux fois plus que les anonymes
- Plus d'un tiers des répondants identifiés ont sélectionné un nombre de fichiers à ouvrir supérieur à 50
  - Principalement des répondants s'étant prononcé au nom de leur institution : Institut Cancer, Fédération Hospitalière de France, Centre Hospitalier de Villeneuve-Saint-George, CH Saintonge...

- Répondants identifiés ayant sélectionné entre 1 et 5 fichiers
- Répondants identifiés ayant sélectionné entre 6 et 50 fichiers
- Répondants identifiés ayant sélectionné entre 51 et 270 fichiers
- Répondants identifiés ayant sélectionné toutes les données listées dans la cartographie (à 3 fichiers près)

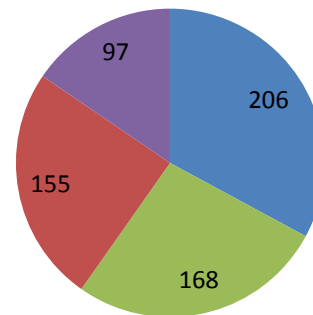
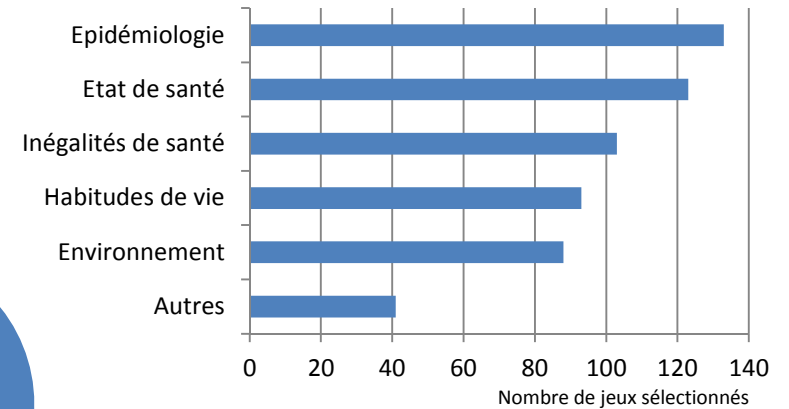


# Les catégories de données les plus demandées

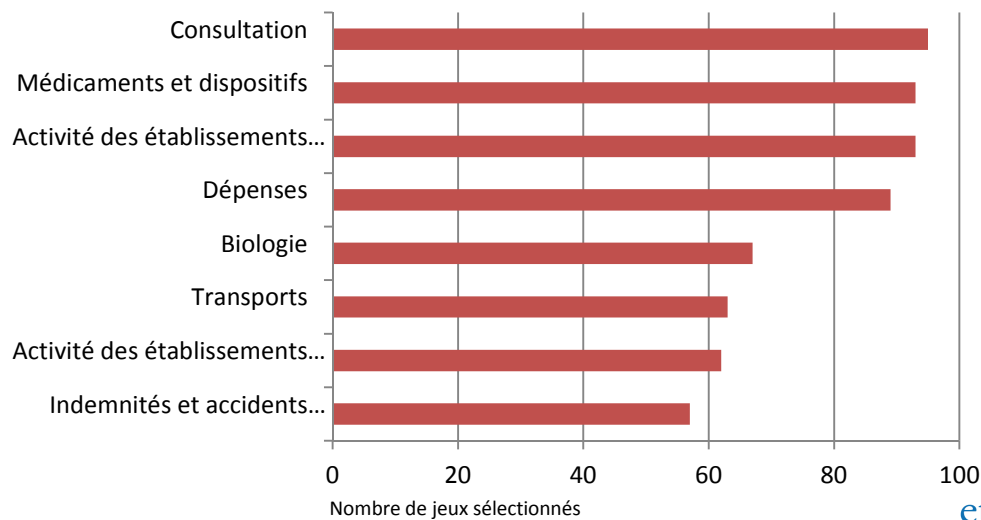
## 4 - Performance et opérations



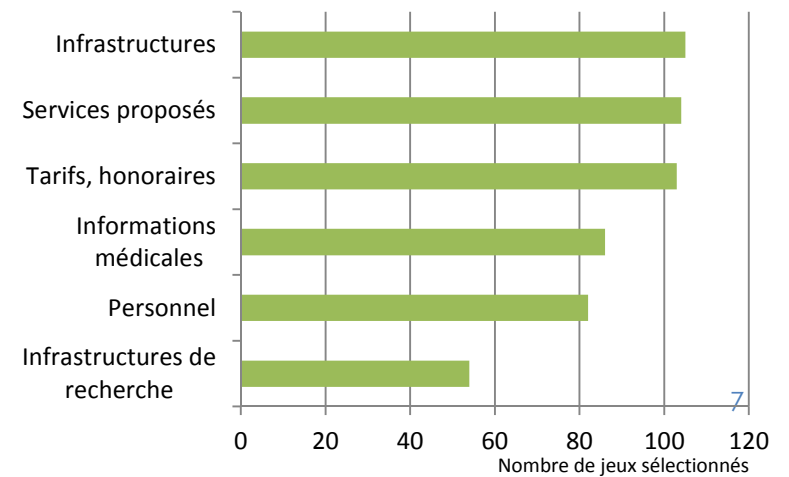
## 1 - Santé publique et épidémiologie



## 3 - Consommation de soins et dépenses



## 2 - Offre de soins



# Les 10 jeux de données les plus demandés

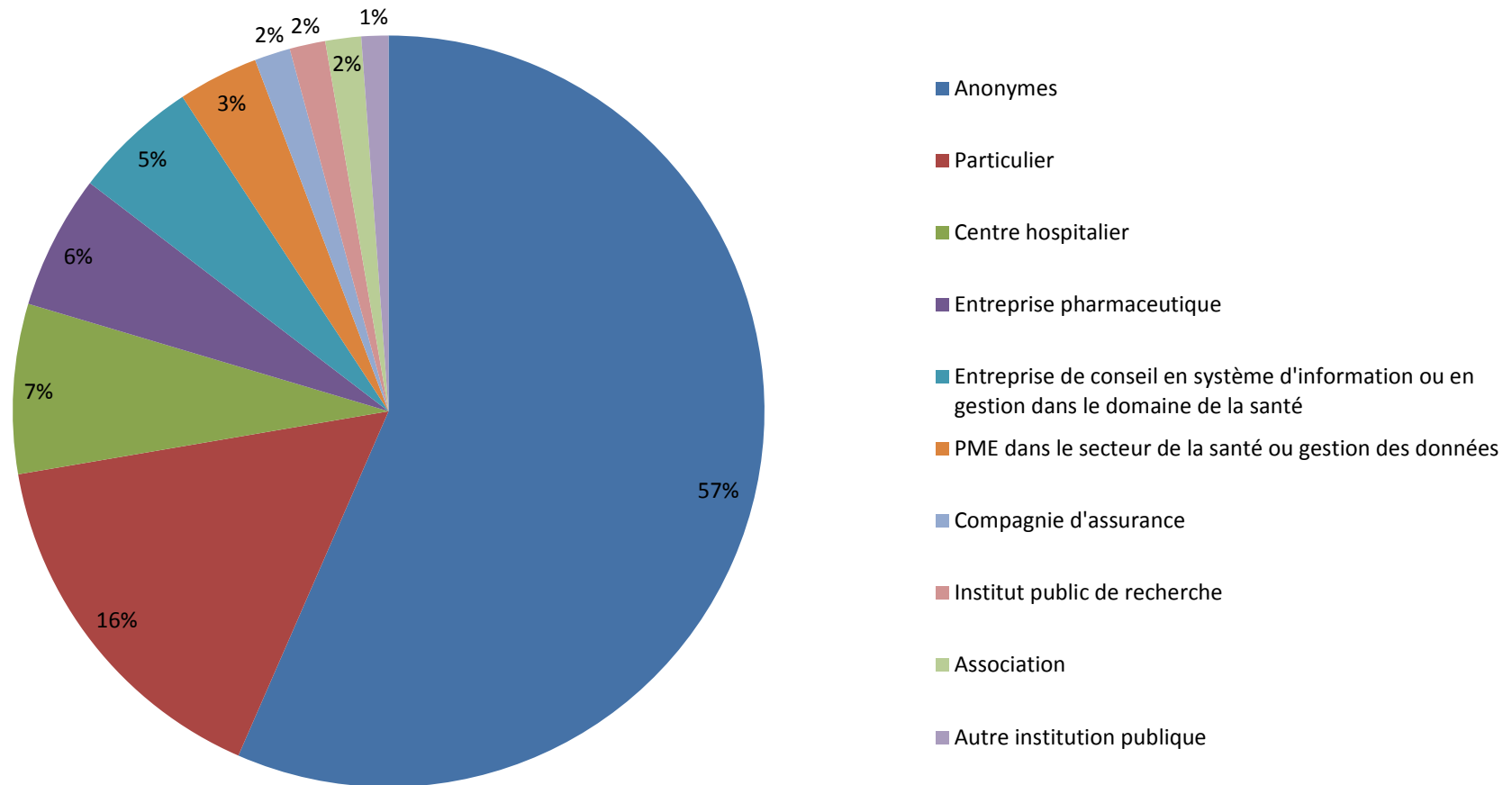
Bases / jeux / enquêtes	Description	Nbre de votes
* <b>Programme de médicalisation des systèmes d'information (PMSI)</b>	Le PMSI contient 4 types de données : 1) Données sur les bénéficiaires 2) Données sur les professionnels de santé ayant prescrit ou réalisé la prestation de soin 3) Données issues des feuilles de soin ou des bordereaux de facturation 4) Données issues des résumés de sortie hospitaliers : dates d'entrée et de sortie, modes d'entrée et de sortie, diagnostic principal, diagnostics associés, actes classants, etc.	262
* <b>Données de consommation inter-régime (DCIR) du SNIIRAM (1)</b>	Le SNIIRAM contient trois types de données 1) Données sur les bénéficiaires 2) Données sur les professionnels de santé ayant prescrit ou réalisé la prestation de soins 3) Données issues des feuilles de soin : nature de la ou des prestations remboursables, date de la ou des prestations, date de la prescription...	176
<b>Base des causes de mortalité</b>	Effectifs de décès et taux bruts de décès par zone géographique, année(s) de décès, causes de décès	55
<b>Tableaux des Affections Longue Durée</b>	- Données sur la fréquence, l'incidence et le coût des ALD - Parmi les données disponibles : par ALD, effectif, taux de prévalence, % d'hommes et de femmes, âge moyen, taux de décès	48
<b>Tableaux "honoraires des professionnels de santé"</b>	Par région et par département : total des honoraires, nombre de dépassements et dépassement moyen, effectifs par tranche d'honoraires	43
<b>Fichiers "analyse des dépenses" (statistiques mensuelles)</b>	Rétrospective annuelle des dépenses d'assurance maladie par catégorie de professionnels de santé et pour chaque catégorie par acte ou par groupe d'actes depuis l'année 2006	43
<b>Données de cartographie du Système national d'information sur l'hospitalisation (SNATIH)</b>	Données d'offre de soin et sur les séjours hospitaliers	41
* <b>Base du Dossier Médical Personnel (DMP)</b>	Données sur les bénéficiaires et les actes réalisés pour les patients ayant un dossier médical de santé	40
* <b>Données de l'échantillon généraliste des bénéficiaires (EGB)</b>	Echantillon permanent représentatif de la population protégée par l'assurance maladie	40
<b>Tableau de bord de l'IDS : suivi des honoraires des médecins par spécialité et lieu d'Exécution</b>	Principaux indicateurs : montant de la dépense remboursable, montant de la dépense remboursée, montant de la dépense présentée au remboursement	38

\* La note suivante accompagnait le questionnaire : « Les jeux de données comportants des données ré-identifiantes sont signalés dans ce questionnaire par une étoile. Ces données ne pourront être mises en open data que sous forme anonymisée et non réidentifiante »



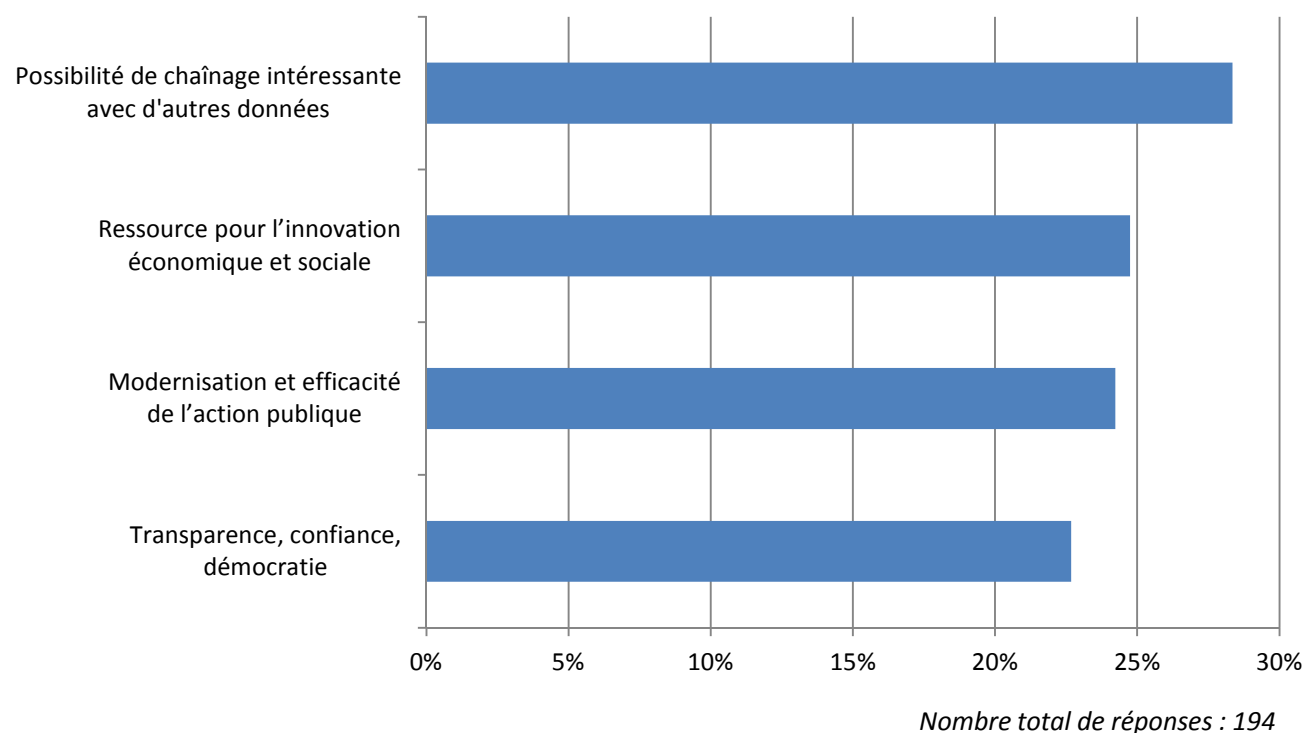
# Le PMSI est de loin la base la plus demandée, avec 262 sélections

Répartition des participants ayant demandé l'ouverture des données agrégées du PMSI



# Le PMSI : Une base « sous utilisée »

Selon vous, l'ouverture de ce jeu de données servira quel(s) objectif(s) :



« les données de la base sont *sous-utilisées* pour la politique publique, c'est une mine d'informations pour la santé et l'économie française ».

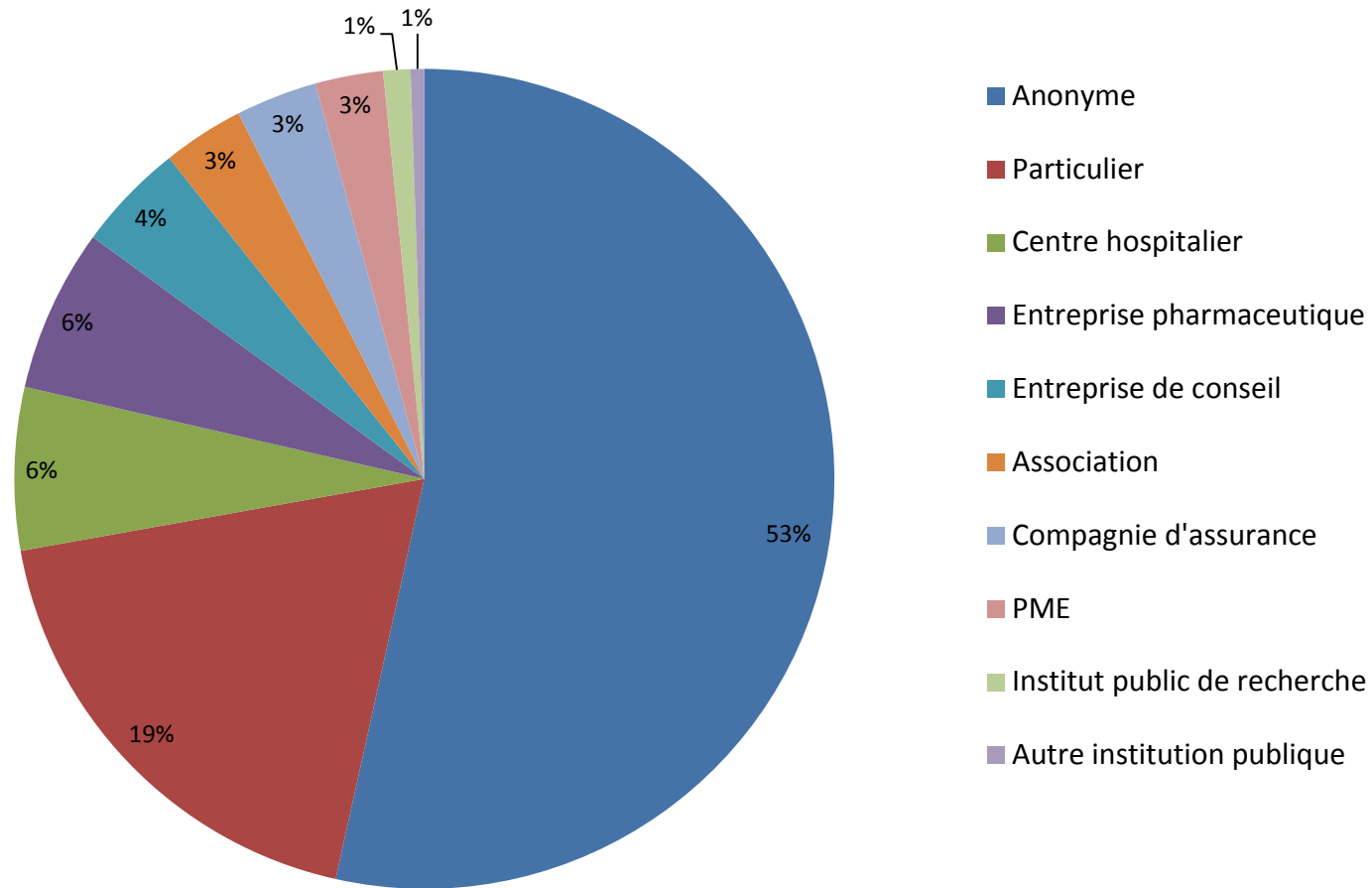
« l'ouverture de cette base servirait les études de suivi dans les pratiques de soins, études épidémiologiques chaînage avec bases où le traitement est indiqué »

« Il est important de publier des données qui respectent l'anonymisation et qui sont géolocalisées à la commune pour faire des corrélations territoriales avec autres jeux de données Open data ». entreprise SOCIO, producteur de logiciel de gestion de données

« Le rapprochement avec les données de consommation de médicaments doit permettre de multiplier l'efficacité de la pharmaco-vigilance », association Pénombre

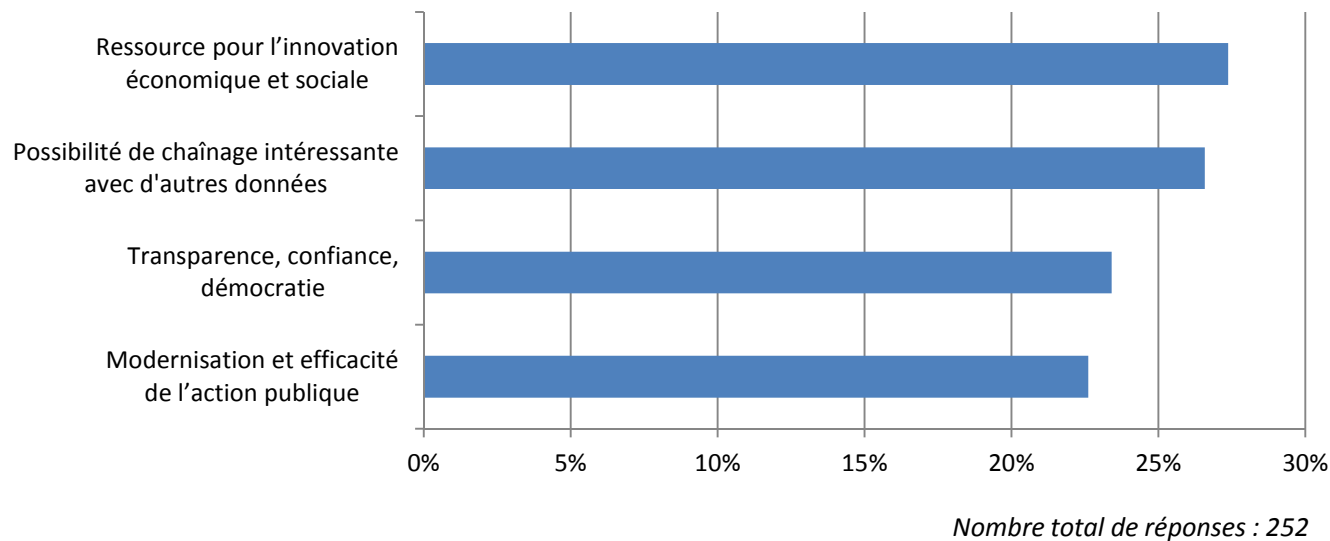
# Les données agrégées du DCIR sont le deuxième jeu de données le plus demandé avec 187 votes.

Répartition des participants ayant demandé l'ouverture des données agrégées du DCIR



# Le DCIR : l'ouverture « pour l'utilisation en recherche clinique »

Selon vous, l'ouverture de ce jeu de données servira quel(s) objectif(s) :

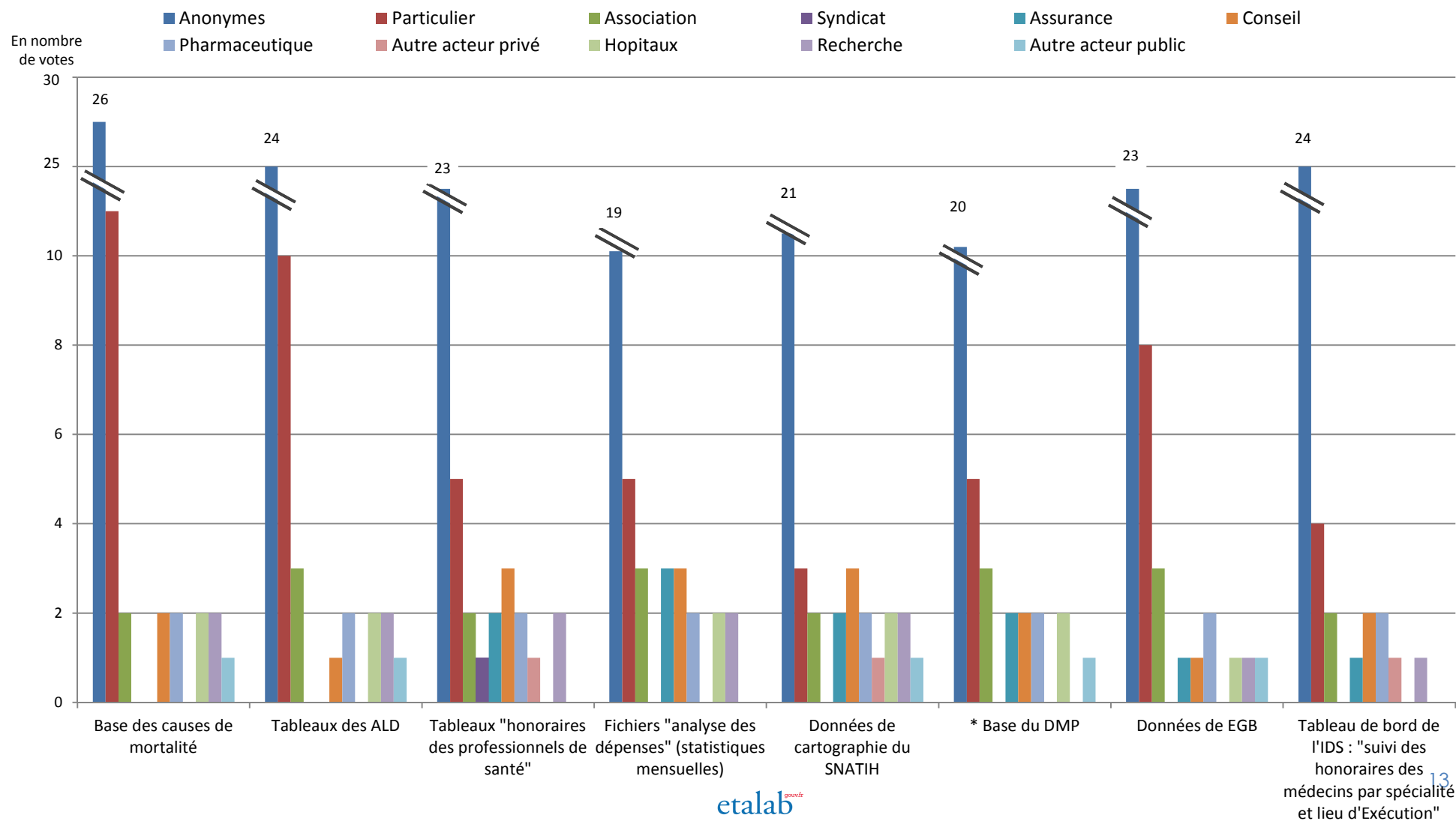


« intérêt pour **utilisation en recherche clinique** »

« les traitements et les actes réalisés en ville pourront permettre **de plus amples études en épidémiologie, évaluations médico-économiques...** »

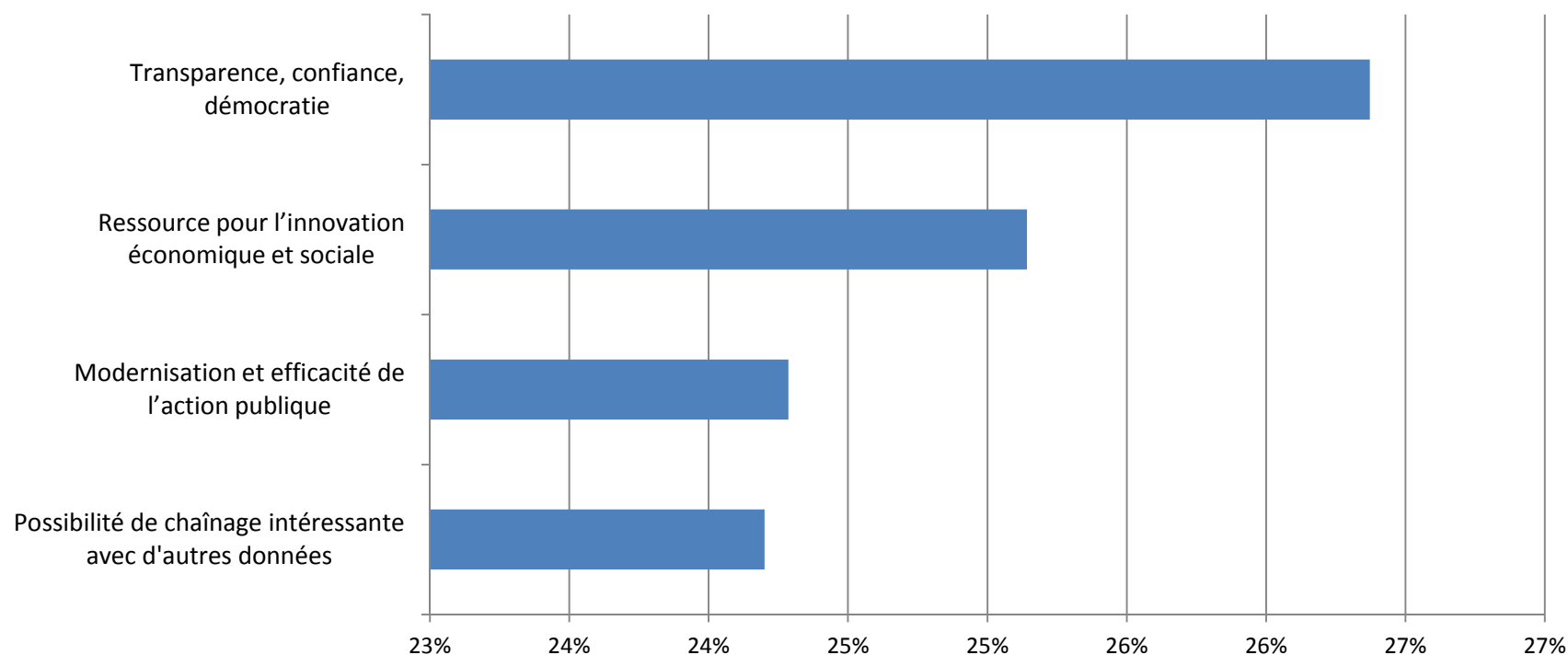
« Le rapprochement avec les données du PMSI doit permettre de **multiplier l'efficacité de la pharmaco-vigilance** »

# Zoom sur les 8 jeux de données les plus sélectionnés, après le PMSI et le DCIR



# La transparence et l'innovation : les deux objectifs principaux de l'ouverture des données de santé

Selon vous, l'ouverture de ce jeu de données servira quel(s) objectif(s) :



Nombre total de réponses : 9 343

# Quelques verbatim

« Les comptes des mutuelles et assureurs, et tous les éléments me **permettant de bien choisir mon assureur complémentaire m'intéressent au plus haut point**, comme la majorité des français, je pense ... »

« Peut on **trouver des liens entre certaines causes de mortalité et des portions de territoires, et donc creuser sur leurs origines potentielle ?** »

« Données de l'enquête sur la santé et la protection sociale : **pourquoi un accès limité à l'IRDES??** »

« L'ouverture de l'échantillon National Témoin Représentatif des personnes Diabétiques (ENTRED) sera intéressante **sous réserve de chaînage possible avec d'autres sources.** »

« Etude Individuelle Nationale de Consommation Alimentaire (INCA) : **faire le lien avec les perturbateurs endocriniens (bisphénol A)** »

« L'ouverture de la base du réseau Sentinelles permettrait **améliorer encore la prévention et l'alerte**, et tenter d'améliorer la propagation »

« **Accès sur le site Ameli: à quand plus de détails des données et pour l'ensemble des codes CIM-10?** »

« L'ouverture des tableaux sur le reste à charge après intervention du régime général d'Assurance Maladie pour l'activité libérale permettra de **réaliser des études d'impact du reste à charge sur l'état de santé, sur le niveau général de consommation de soins etc.** »

« Base de données sur la qualité de l'air (BDQA) : **faire le lien avec les AVC** »

« Base de l'établissement français du sang (EFS) : **la proportion des personnes qui ne peuvent pas donner leur sang doit être indiqué** à mon sens car cela peut être dû à l'existence d'un problème de santé. »

« L'ouverture en open data du RPPS, le RASS et les RoR permettrait de **s'assurer de la fiabilité des données fournies par un acteur**, de son existence »

# Les participants ont mentionné certains indicateurs qu'ils recherchaient ou qu'ils proposaient de créer

- Indicateur de satisfaction des soignants intra établissement
- Habitude alimentaire en fonction de l'activité/ le métier, la dépense calorifique
- Apports nutritionnels complets pour chaque produit du marché dans le but d'indiquer au consommateur, selon ses carences, quel produit consommer, en telle quantité, à et tel moment.
- Modèles mathématiques reliant l'IMC, l'âge, le sexe, l'ethnie, l'activité au sens large (physique, intellectuelle, etc.)
- Qualité de l'alimentation dans les hôpitaux
- Dans un objectif de santé publique, il faudrait publier pour chaque pathologie (suivre la codification CIM10):- des descriptions des patients atteints : caractéristiques démographiques- les estimations en termes de prévalence et d'incidence- les prises en charges : lignes de traitement, actes médicaux, examens complémentaires, hospitalisations, transports- la qualité de vie de ces patients
- Dans la base de l'ANSM on ne trouve pas les dispositifs liés aux pathologies. Par exemple, on y trouve les insulines, mais aucune mention des lancettes, aiguilles, bande de test qui sont pourtant remboursés par la sécurité social.
- Informations structurées sur les interactions médicamenteuses. Elles ne sont publiées par la HAS que sous forme de PDF Sécurité du patient
- Indicateur de la production des déchets et leur récupération (cela donne une indication sur l'importance ou non de personnes atteintes d'une maladie) par département et région
- Indicateur sur la quantité de produits phytosanitaires achetée (ou utilisée) dans un département
- Indicateur sur le nombre de plaintes d'ordre sanitaire (plainte bruit...)
- Indicateurs sur le reste à charge des assurés après prise en charge par l'AMO et l'AMC
- Profils médecins générateurs d'arrêt de travail