

Indicateurs de mesure de la qualité du RSE

Dominique Francoz
INSEE



Mesurer pour comprendre



09/07/17

Les différentes composantes de la qualité

◆ Pertinence :

- Dans quelle mesure le RSE satisfait les besoins des utilisateurs actuels et potentiels

◆ Précision :

- Dans quelle mesure l'information enregistrée correspond à la réalité

◆ Ponctualité :

- Les données fournissent une image du monde réel avec le plus petit laps de temps

◆ Comparabilité :

- Comparabilité temporelle et spatiale

Les différentes composantes de la qualité

◆ Accessibilité :

- Possibilité pour les utilisateurs internes d'obtenir des données individuelles en se connectant directement à la base de données
- Possibilité pour les utilisateurs externes d'obtenir des données agrégées

◆ Cohérence

- Cohérence interne au RSE et avec les autres répertoires

◆ Coût et efficacité

- Le coût inclut la charge de réponse des entreprises

Mesurer la qualité d'un RSE

- ◆ **Différent de la mesure de la qualité d'un processus d'enquête**
 - Multiplicité des sources
 - L'INS n'a pas toujours la maîtrise des informations
 - Hétérogénéité des utilisateurs
- ◆ **Utilisation des sources administratives :**
 - La qualité du RSE est fortement liée à la qualité de la source administrative
 - L'INS n'a pas le contrôle sur cette qualité
 - Définitions, caractéristiques ⇒ **impact sur la précision**

Mesurer la qualité d'un RSE

◆ Hétérogénéité des inputs

- La construction d'un RSE fait généralement appel à plusieurs sources d'informations administratives ou statistiques
- Chaque source fournit une information partielle ou pour une sous-population ⇒ **impact sur la cohérence**

◆ Variabilité des inputs dans le temps

- L'administration peut changer ses règles de gestion sans concertation avec l'INS, changement des règles d'enregistrement ⇒ **impact sur la comparabilité temporelle**

Mesurer la qualité d'un RSE

◆ Le produit principal est la micro-donnée

- On ne peut pas faire de traitements comme sur les données agrégées
- Les erreurs de sous-couverture et de sur-couverture ne se compensent pas

◆ Hétérogénéité des utilisateurs

● Besoins contradictoires

- Lancement d'enquêtes : ponctualité
- Enquêtes sur le chiffre d'affaires : focus sur les grandes entreprises ⇒ présence des grandes entreprises importante
- Démographie des entreprises : focus sur les petites entreprises ⇒ exhaustivité importante

Mesurer la qualité d'un RSE

◆ Hétérogénéité des utilisateurs

● Besoins contradictoires

- Variétés des utilisateurs externes : veulent des données fiables sur leur domaine d'investigation ⇒ en contradiction avec l'utilisation d'un répertoire unique pour tirer les échantillons

Traitement de la qualité d'un RSE

◆ En trois étapes

- **Qualité des inputs (sources administratives)**
 - Grille d'analyse de la qualité des sources administratives en amont de leur utilisation
- **Qualité du processus de création/mise à jour du répertoire**
- **Qualité des outputs**

Qualité des données administratives

◆ Pertinence

Indicateur qualitatif	Description
Nommer chaque source administrative utilisée comme input dans le RSE et décrire l'objectif principal de cette collecte de données	Nommer toutes les sources administratives et leurs producteurs. Cette information va aider les utilisateurs à évaluer si le produit statistique est pertinent pour l'usage envisagé (ici création et mise à jour du RSE) L'information sur l'objectif principal de la collecte aide aussi les utilisateurs à décider si la source est pertinente pour leurs besoins.

Qualité des données administratives

◆ Pertinence

Indicateur qualitatif	Description
Décrire l'utilisation principale de chaque source administrative et, quand c'est possible, comment les données sont en adéquation avec les besoins des utilisateurs	Inclure les principaux processus statistiques connus pour réclamer des données de sources administratives Cet indicateur devrait aussi saisir dans quelle mesure les données soutiennent les besoins des utilisateurs ; Cette information peut être obtenue via des enquêtes de satisfaction des utilisateurs.

Qualité des données administratives

◆ Pertinence

Indicateur qualitatif

Description

Décrire dans quelle mesure les données de la source administrative répondent aux besoins statistiques

Les besoins statistiques sont à définir et le degré d'adéquation de la source administrative avec ces besoins évalués. Les écarts entre les données administratives et les besoins statistiques peuvent avoir un effet sur la pertinence pour l'utilisateur. Tous les écarts et raisons de la non-exhaustivité d'une source doivent être décrits, par exemple si certaines zones de la population cibles ne sont pas couvertes ou si certaines variables qui pourraient être utiles ne sont pas collectées. Toute méthode pour remplir les trous devrait être stipulée.

Qualité des données administratives

◆ Pertinence

Indicateur qualitatif	Description
Décrire les contraintes liées à la disponibilité des sources administratives au niveau de détail requis	Certaines micro-données administratives ont une disponibilité limitée ou peuvent être disponibles seulement à un niveau agrégé ; Décrire les restrictions sur le niveau des données disponibles et l'effet sur le produit statistique.

Qualité des données administratives

◆ Pertinence

Indicateur qualitatif	Description
Identifier les trous entre les besoins des utilisateurs clés en termes de couverture et de détail et les données courantes	Les données sont considérées comme complètes quand elles satisfont les besoins des utilisateurs en termes de couverture et de détail. Cet indicateur permet aux utilisateurs d'évaluer, en cas de trous, comment les outputs sont pertinents au regard de leurs besoins

Qualité des données administratives

◆ Ponctualité et fraîcheur

Indicateur qualitatif	Description
Décrire l'intervalle de temps entre deux livraisons de données administratives	
Décrire dans quelle mesure les données administratives sont fraîches	Décrire combien de temps s'écoule entre la production de la donnée administrative et sa livraison à l'institut statistique. Décrire les effets d'un manque de ponctualité

Qualité des données administratives

◆ Ponctualité et fraîcheur

Indicateur qualitatif	Description
Décrire tout défaut de ponctualité dans la livraison des données administratives	Donner le détail de l'écart entre la date de livraison prévue et la date effective. Toute raison pour cet écart doit être documentée ainsi que les effets sur le produit statistique
Fréquence de production	Est-ce que la fréquence de mise à jour du RSE correspond aux besoins des utilisateurs ?

Qualité des données administratives

◆ Cohérence

Indicateur qualitatif	Description
Décrire les identifiants communs des unités de la population dans les sources administratives	Différentes sources administratives ont souvent des identifiants différents. Quand on est en présence de plusieurs sources, on a souvent besoin d'apparier les enregistrements de plusieurs sources pour une même unité. La présence d'un identifiant commun facilite cet appariement.
Décrire la marge de tolérance et ses raisons d'être	Quand on utilise plusieurs sources administratives qui observent les mêmes variables pour les mêmes unités, on peut observer des incohérences entre les sources. L'amplitude maximale d'incohérence doit être fixée ainsi que le traitement proposé

Qualité des données administratives

◆ Cohérence

Indicateur qualitatif	Description
Décrire les différences de concept, définitions et classifications entre la source administrative et le résultat	Les concepts incluent la population, les unités, les domaines, les variables et la référence temporelle qui peuvent être différents dans la source administrative et dans le résultat final. Les problèmes de référence temporelle se produisent quand l'INS demande des données pour une certaine période de référence et obtient des données pour une autre période de référence. Il s'agit d'identifier les effets de ces différences et décrire les moyens qui sont pris pour y remédier.

Qualité des données administratives

◆ Cohérence

Indicateur qualitatif	Description
Décrire les ajustements faits pour les différences de concepts et définitions entre la source administrative et le résultat attendu	Décrire pourquoi un ajustement est nécessaire et comment il est opéré. Cette information doit être fournie aux utilisateurs afin qu'ils puissent comprendre d'éventuelles différences entre le résultat statistique et les informations provenant d'autres sources de données.

Qualité des données administratives

◆ Précision

Indicateur qualitatif	Description
Décrire les procédures d'appariement réalisées entre sources administratives	Décrire la technique d'appariement utilisée (automatique vs. Manuel) et son taux de succès.
Décrire le processus requis pour traiter la non-réponse dans la source administrative	

Qualité des données administratives

◆ Précision

Indicateur qualitatif	Description
Décrire les problèmes d'erreurs de classification et comment on les traite	Décrire les différences entre les classifications utilisées dans les sources administratives et celles requises pour le résultat statistique. Décrire, s'il y a lieu, les procédures de redressement.
Décrire la couverture de la source administrative et les problèmes éventuels de couverture	L'idée est d'évaluer si la couverture est suffisante ou non. On peut identifier des problèmes de sous-couverture(enregistrements manquants) ou de sur-couverture (duplications d'enregistrements).

Qualité des données administratives

◆ Précision

Indicateur qualitatif	Description
Décrire les traitements préalables à l'utilisation de la source administrative	Des traitements sont parfois nécessaires pour vérifier ou améliorer la qualité des données administratives ou créer des variables nouvelles pour des besoins statistiques.

Qualité des processus

◆ Intégration des sources administratives

- Nombre d'unités dupliquées en pourcentage du nombre total d'unités
- Nombre de nouveaux enregistrements durant une période de temps donnée

Qualité des output

◆ ponctualité

- Laps de temps pour chaque enquête entre la période de référence de l'enquête et la date de livraison de la base de sondage

◆ Précision

- Couverture: complétude de la base de sondage à la période de référence de l'enquête
- Complétude des caractéristiques
- Taux d'erreur (% d'unités avec des adresses erronées, % d'unités avec des codes d'activité erronés...)
- Taux de non-réponse (indicateur d'erreur dans les informations de contact)

◆ Accessibilité: dans quelle mesure les données pertinentes sont disponibles pour les utilisateurs

◆ Cohérence: avec les autres répertoires

Des indicateurs de qualité

- ◆ **Nombre d'entreprises sans code d'activité valide**
- ◆ **Estimation du nombre d'entreprises mal classifiées**
 - **Activité économique**
 - **Localisation**
 - **taille**
- ◆ **Nombre d'entreprises pour lesquelles le statut d'activité est incertain**
 - **Nombre d'entreprises ayant un statut d'activité « actives » mais qui sont en réalité cessées ou inactives**
- ◆ **Evaluation de la sous-couverture**
 - **Dûe à la sous-couverture de la source administrative**
 - **Dûe au délai d'enregistrement des nouvelles entreprises**
- ◆ **Estimation de l'incidence de mauvaises information de contact**

.....pour la mise en place d'une politique d'amélioration de la qualité

◆ Calculer les indicateurs de qualité

◆ Fixer des objectifs

- Diminuer de 25% le nombre d'unités sans code d'activité valide
- Multiplier par deux le nombre d'enquêtes qui utilisent le répertoire comme base de sondage
- ◆ Mise en place de mesures pour améliorer le répertoire
 - Améliorer la complétude en investiguant de nouvelles sources potentielles de mise à jour
 - Enquêtes d'amélioration du répertoire ciblant des populations particulières (unités potentiellement inactives, unités sans code d'activité ou ayant un code potentiellement faux....)

Merci de votre attention

Avez-vous des questions ?



Dominique FRANCOZ

Division ATI

+33141175301

Dominique.francoz@insee.fr

Insee

Direction Générale

www.insee.fr

 [@InseeFr](https://twitter.com/InseeFr)