

# ***Sources et méthodes***

## **L'enquête sur les technologies de l'information et de la communication, auprès des entreprises**

L'enquête sur les technologies de l'information et de la communication (TIC) a pour objectif de mieux percevoir la diffusion des TIC dans les entreprises. Elle s'inscrit dans le cadre du règlement européen concernant les statistiques communautaires sur la société de l'information, qui exige la fourniture de données jusqu'en 2019, afin d'évaluer la progression de l'usage des TIC dans les entreprises européennes.

---

## **Présentation de l'enquête TIC**

---

Le programme d'enquêtes sur les technologies de l'information et de la communication (TIC) a pour objectif de mieux percevoir la diffusion des TIC dans les entreprises. Il cherche notamment à apprécier la place des outils nouveaux dans les relations externes de l'entreprise (Internet, commerce électronique) et dans leur fonctionnement interne (réseaux, systèmes intégrés de gestion). Un des objectifs de l'enquête est de faire le point sur l'adoption et l'utilisation des technologies de l'information et de la communication par les entreprises françaises, et d'en analyser les évolutions. Cette enquête sert de support aux réponses françaises transmises à Eurostat.

Ce programme d'enquêtes s'inscrit dans le cadre du règlement n°808/2004 du 21 avril 2004, amendé par le règlement européen n° 1006/2009 du 16 septembre 2009, du Parlement européen et du Conseil concernant les statistiques communautaires sur la société de l'information et vise à satisfaire les besoins d'informations tant au niveau européen qu'au niveau national.

Le programme d'enquêtes sur les TIC se compose d'une séquence d'enquêtes annuelles conçues autour d'un tronc commun portant sur trois thèmes principaux (équipement en TIC, accès et usage d'Internet, commerce électronique) auquel est adjoint chaque année un module traitant d'un thème nouveau ou approfondissant l'un des thèmes de base (par exemple, l'utilisation des technologies basées sur l'Identification par Radio Fréquence - RFID).

Le tronc commun du questionnaire se compose d'une cinquantaine de rubriques, le module annuel d'une quinzaine.

---

## **La méthodologie de l'enquête**

---

### **Le champ de l'enquête TIC**

L'unité d'interrogation est l'entreprise-unité légale.

Pour se rapprocher des concepts utilisés dans l'enquête européenne, dans l'enquête TIC 2011, la taille des unités interrogées n'est plus mesurée en nombre de salariés, mais en nombre de personnes occupées : les personnes non-salariées sont incluses dans les effectifs. L'enquête Tic 2011 porte ainsi sur les sociétés d'au moins 10 personnes occupées, - ce qui comprend notamment les unités employant 9 salariés et dont le dirigeant est non salarié - les enquêtes précédentes sur celles d'au moins 10 salariés. L'une des conséquences de cette modification sur les résultats est une augmentation de l'influence des plus petites sociétés.

Le champ recouvre l'ensemble des entreprises-unités légales marchandes exploitantes françaises (de la métropole pour les années 2008 à 2010) appartenant :

1. Jusqu'à l'enquête TIC 2008, aux secteurs de la NAF rév. 1 suivants :

- l'industrie et les industries agroalimentaires (section D) ;
- le secteur de l'énergie (section E) ;
- la construction (section F) ;
- le commerce (section G) ;
- les hôtels et restaurants (section H) ;
- les transports et communications (section I) ;
- les services financiers (section J partielle) ;
- l'immobilier, la location et les services aux entreprises (section K) ;
- le secteur de l'audiovisuel (groupes 92.1 et 92.2).

2. A partir de l'enquête TIC 2009, aux secteurs suivants de la NAF rév. 2 (entrée en vigueur au 1<sup>er</sup> janvier 2008) :

- l'industrie manufacturière (section C) ;
- la production et la distribution d'électricité, de gaz, de vapeur et d'air conditionné (section D) ;
- la production et la distribution d'eau, l'assainissement, la gestion des déchets et la dépollution (section E) ;
- la construction (section F) ;
- le commerce, la réparation d'automobiles et de motocycles (section G) ;
- les transports et l'entreposage (section H) ;
- l'hébergement et la restauration (section I) ;
- l'information et la communication (section J) ;
- les activités immobilières (section L) ;
- les activités spécialisées, scientifiques et techniques hors activités vétérinaires (divisions 69 à 74) ;
- les activités de services administratifs et de soutien (section N) ;
- la réparation d'ordinateurs et d'équipements de communication (groupe 95.1).

L'enquête couvre également pour les enquêtes TIC 2009 et TIC 2010 :

- les autres intermédiations monétaires et les autres distributions de crédit (classes 64.19 et 64.92) ;
- les assurances et la réassurance (groupes 65.1 et 65.2) ;
- les activités auxiliaires de services financiers, hors assurance, les caisses de retraite et l'administration de marchés financiers (classes 66.12 et 66.19).

3 - A partir de l'enquête 2011, les sociétés financières et d'assurance sont exclues. Les secteurs d'activité suivants ne font plus partie du champ de l'enquête, contrairement aux années précédentes :

- les autres intermédiations monétaires et les autres distributions de crédit (classes 64.19 et 64.92 de la NAF rév. 2) ;
- l'assurance et la réassurance (groupes 65.1 et 65.2) ;
- l'administration de marchés financiers et les autres activités auxiliaires de services financiers, hors assurance et caisses de retraite (classes 66.11 et 66.19).

## Le plan de sondage de l'enquête

L'échantillon de l'enquête TIC est tiré dans une base de sondage construite à partir du [répertoire Sirene](#). La méthode d'échantillonnage est un sondage aléatoire stratifié selon le secteur d'activité, les tranches d'effectifs et depuis 2012, le chiffre d'affaires. La stratification adoptée est donc un croisement entre le secteur d'activité de l'unité, sa tranche d'effectif (et un seuil de chiffres d'affaires depuis 2012).

Les modalités des secteurs d'activité ont des niveaux d'agrégation très divers (de la classe au regroupement de sections), en NAF rév. 1 jusqu'en 2008, comme en NAF rév. 2 à partir de l'enquête de 2009. Mais il peut y avoir des regroupements de divisions ou de groupes.

Les modalités des tranches d'effectifs sont au nombre de cinq (10 à 19 personnes occupées, 20 à 49, 50 à 249, 250 à 499, 500 et plus).

Les seuils de chiffre d'affaires (CA) sont différents selon la tranche d'effectif :

- 10 à 19 personnes occupées : CA  $\geq$  50 millions d'euros
- 20 à 49 personnes occupées : CA  $\geq$  200 millions d'euros

- 50 à 249 personnes occupées : CA  $\geq$  800 millions d'euros
- 250 à 499 personnes occupées : CA  $\geq$  1 500 millions d'euros
- 500 personnes occupées et plus : pas de distinction selon le CA car strates exhaustives.

Pour les tranches d'effectifs jusqu'à 249 personnes, ces seuils ont été déterminés à partir du chiffre d'affaires collecté dans l'enquête TIC 2009. Le principe est de retenir le chiffre d'affaires à partir duquel une unité pondérée par le poids de sondage moyen de sa tranche d'effectif représente plus de 1% du CA global de sa tranche d'effectif, estimé par l'enquête. Pour plus de lisibilité, ce CA a été arrondi. Pour la tranche 250 à 499 personnes, le seuil a été fixé à 1 500 afin de correspondre à celui de la définition des grandes entreprises dans le cas où l'effectif est inférieur à 5000.

Le nombre d'unités à échantillonner diffère selon les strates. Les unités de plus de 500 sont interrogées exhaustivement. Pour les strates de taille d'effectif inférieure, le nombre d'unités à interroger a été obtenu par le biais d'une allocation proportionnelle à l'effectif.

*In fine*, selon l'année d'enquête, environ 13 000 unités sont interrogées.

## Le traitement de la non-réponse et le calage

Comme pour toutes les enquêtes, les résultats bruts de l'enquête TIC sont traités pour répondre à deux objectifs principaux :

1. corriger le biais introduit par les non-réponses totales et partielles ;
2. réduire autant que possible les fluctuations d'échantillonnage résultant du redressement de la non-réponse.

Les résultats de l'enquête TIC seraient biaisés si l'on ne corrigeait pas la non-réponse, sauf dans l'hypothèse où les non-répondants ont un comportement identique à celui des répondants. Mais, les non-répondants ne se répartissent *a priori* pas au hasard. Aussi, afin de compenser le biais lié à la non-réponse totale, on effectue les traitements suivants.

### Traitements préalables

Les unités enquêtées sont ventilées dans une des 4 catégories suivantes :

- a) **Les unités répondantes et qui appartiennent au champ de l'enquête ;**
- b) **Les unités du champ reconnues comme non-répondantes:** Cette catégorie englobe les unités n'ayant pas retourné de questionnaire, et celles ayant renvoyé un questionnaire inexploitable ;
- c) **Les unités repérées comme hors-champ ;**
- d) **Les unités sans information :** il s'agit des unités pour lesquelles aucun questionnaire n'a été réceptionné, et dont on ne sait donc pas si elles sont hors-champ ou non-répondantes totales.

Une des premières étapes du redressement de la non-réponse consiste à ventiler les unités sans information dans les autres catégories. Pour ce faire, des sources externes sont utilisées : SIRENE et TVA. Si ces recherches effectuées via ces sources externes ne permettent pas de statuer, l'unité sera alors basculée par défaut dans le groupe des unités non-répondantes. Au final, on n'a donc plus que 3 catégories.

### Redressement de la non-réponse partielle

Le redressement de la non-réponse partielle concerne les unités répondantes du champ. Cette étape permet de compléter les quelques items auxquels l'enquêté n'a pas répondu.

Après avoir mis en place certains automatismes de réponse pour les unités répondantes du champ (règles déterministes, par exemple : « si j'ai déclaré avoir accès à internet, alors je suis supposé avoir un ordinateur » etc...), on répartit chaque variable du questionnaire de ces unités, selon trois situations :

- l'unité **n'est pas concernée** par la question et n'a donc pas à y répondre (exemple : « si je n'ai pas d'ordinateur, je ne réponds pas aux questions sur l'internet ») ;
- l'unité est concernée par la question, et y **a répondu** ;
- l'unité est concernée par la question, mais n'y a **pas répondu**.

Ce sont uniquement les variables de ce dernier cas qui vont être redressées dans le cadre de la correction de la non-réponse partielle. Après traitement, elles présenteront une réponse.

De façon schématique : on utilise les renseignements donnés par les unités répondantes d'une catégorie (critères de taille, d'activité, de comportement, etc.), pour compléter les questions manquantes des unités répondantes (partielles) de la même catégorie.

On distingue, lors des traitements, les variables qualitatives des variables quantitatives.

Pour les variables qualitatives, on procède à une imputation aléatoire. L'imputation est faite en simulant la distribution des réponses observées parmi les unités ayant les mêmes caractéristiques. Les caractéristiques prises en compte peuvent être des caractéristiques de l'unité, connues avant l'enquête ou bien aussi des réponses à d'autres questions. Ceci permet de conserver un lien statistique entre certaines variables. L'ordre des traitements a une importance capitale.

Quant aux variables quantitatives, le redressement se porte dans un premier temps sur les items en pourcentage (les réponses peuvent être exprimées en montant ou en pourcentage) Ainsi, lorsque l'unité a renseigné le montant et non le pourcentage, une règle de trois est appliquée. De même, après le redressement des pourcentages, une règle de trois est utilisée afin d'imputer l'item en valeur.

Dans le cas de l'imputation des pourcentages, le redressement s'effectue de façon similaire à celui des variables qualitatives. En effet l'analyse des réponses montre clairement que les unités déclarent un ordre de grandeur pour les pourcentages et non une réponse « exacte ». De fait peu de modalités sont citées (0, 5, 10, les multiples de 10 et 95). La réponse en pourcentage est donc convertie en variable qualitative, en effectuant des regroupements par classe. Cette variable qualitative est redressée classiquement par imputation aléatoire. Ensuite cette variable qualitative est transformée en variable quantitative en prenant la réponse médiane de la tranche.

#### Cas des unités non substituables

Les unités dites non-substituables correspondent aux unités très grandes et/ou très particulières par rapport aux thèmes traités.

Le redressement de la non-réponse partielle de ces unités a distingué :

- les unités non substituables répondantes (partielles) du champ ;
- les unités non substituables non-répondantes (totales) du champ, pour lesquelles la non-réponse totale a été redressée comme de la non-réponse partielle.

Pour les unités non substituables renvoyant leur questionnaire, mais ne répondant que partiellement, les données de l'enquête précédente sont utilisées, lorsque cela est possible. Ensuite, la non-réponse partielle qui subsiste sur ces unités est redressée de façon similaire à la non-réponse partielle des unités substituables.

### **Redressement de la non-réponse totale et calage**

Rappel : Les unités « non-substituables » ainsi que les unités non-répondantes totales basculées en non-réponse partielle ont été exclues des traitements de la non-réponse totale.

La méthode de redressement de la non-réponse totale utilisée consiste à bâtir un modèle explicatif du comportement de réponse, et à retenir les variables qui apparaissent significatives. Les variables retenues doivent être connues sur les unités répondantes et sur les autres, ce qui exclut le recours aux questions de l'enquête.

Des groupes de réponses homogènes (GRH) sont constitués à l'aide des caractéristiques disponibles à la fois sur les répondants et sur les non-répondants, et qui se révèlent avoir une influence significative sur la probabilité de réponse.

On corrige la non-réponse totale, en « transférant » le poids initial des unités non-répondantes totales du champ, sur les autres unités de l'enquête, en fonction de leur GRH. Cela permet de prendre en considération les caractéristiques des unités et les probabilités de réponse de chaque groupe.

Le poids des unités « non-substituables » ainsi que des unités « basculées en non-réponse partielle » n'a pas été modifié.

La dernière étape du traitement est le calage. En effet, après re-répartition des poids comme décrit ci-dessus, la somme des poids par strate n'est plus identique à la somme des poids initiaux par strate.

Ces écarts peuvent être redressés par une modification des poids des individus de l'échantillon. Cette repondération est réalisée à l'aide d'une méthode usuelle de « calage sur marges ». La méthode consiste à modifier les poids des unités de telle sorte que le nombre d'unités total, estimé à partir des répondantes de l'échantillon, soit égal au total (ou « marges ») connu par ailleurs, en l'occurrence ici, le nombre d'unités dans le champ de l'enquête.

L'idée générale du calage est d'obtenir, en utilisant les poids issus du calage, un nombre d'unités par strate de diffusion égal à celui présent dans la population au départ. Le calage s'effectue en ne tenant compte ni des unités non-substituables, ni des unités non-répondantes totales passées en non-répondantes partielles. Ces unités garde leur poids initial (de lancement). Les unités déclarées hors-champ participent au calage.

L'échantillon obtenu, résultant de la collecte, est calé sur les données provenant du répertoire Sirene : les marges sont le nombre d'unités issu de Sirene selon les strates de diffusion au sens d'Eurostat.