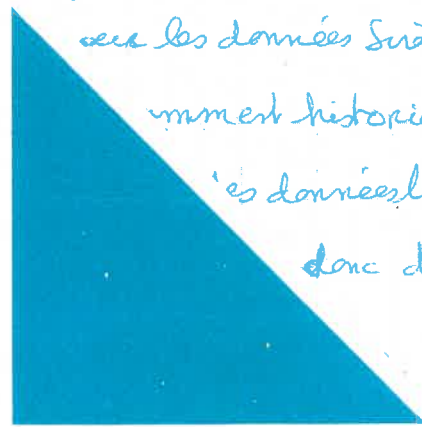


Note et appréciations du correcteur :

N.B. - Il est interdit aux candidats de signer leur composition ou d'y mettre un signe quelconque pouvant indiquer la provenance de la copie

et fortement impacté. Les
sur les données Suède,
mment historiques,
es données locale
donc d'élarg



ÉPREUVE

de Note de synthèse (administratif)

NOMBRE D'INTERCALAIRES : 2

Dès 2003 avec l'essor d'Internet, l'Insee a mis en place une politique d'ouverture des données, l'open data.

Cela consistait essentiellement à la mise en ligne gratuite de publications et de fichiers de données sur son site internet.

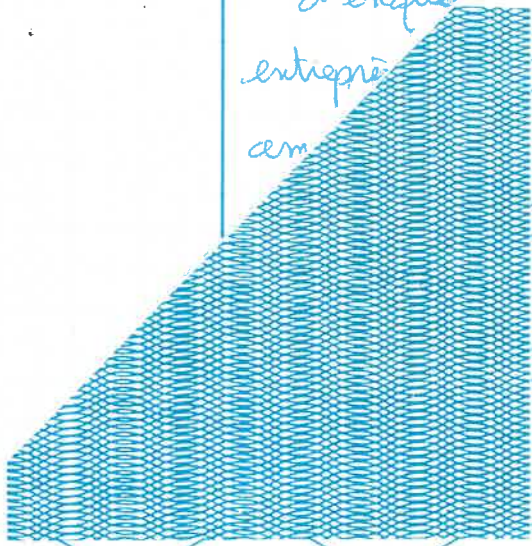
Cependant, les modalités de cette offre sont modifiées par l'adoption de la loi n° 2016-1321 du 7 octobre 2016, dite loi pour une République numérique.

En effet, cette loi stipule entre autres la modernisation des échanges de données dans le cadre de la statistique publique.

Dans ce contexte, quelles sont les conséquences concrètes de la loi pour une République numérique sur le service statistique public français (SSP)?

Dans un premier temps, sera étudié l'impact de cette loi sur l'accès aux différentes données concernées (I). Dans un second temps, seront abordés les divers traitements facilités par cette loi (II).

d'enquêt
entrepr
am



I - Un accès renforcé aux différentes données.

La loi pour une République Numérique renforce à la fois la diffusion des données détenues par l'Insee (A) et celles transmises par des entreprises privées (B).

A - Une diffusion élargie des données Sirene

Selon la loi pour une République numérique, la base Sirene doit être rendue disponible gratuitement pour tout utilisateur public ou privé et accessible par voie informatique. Cela entraîne plusieurs conséquences.

Tout d'abord, cette gratuité des données a un impact financier ^{fort} puisque les recettes liées à la diffusion Sirene étaient estimées à 11 millions d'euro par an pour l'Insee. Néanmoins, la direction du Budget s'est engagée à compenser cette perte.

Ensuite, la diffusion des données est fortement impactée. Les fichiers Syracuse, base jusqu'ici diffusée pour les données Sirene, ne sont pas exhaustifs. D'autres données, notamment historiques, peuvent compléter le dispositif, Des séries longues, des données locales et certains indicateurs sont aussi concernés. Il s'agit donc d'enrichir l'offre tout en garantissant une diffusion de qualité.

Enfin, le dernier fort impact de cette loi sur les données Sirene est en effet son mode de diffusion. La gratuité des données et leur élargissement pourrait entraîner une forte hausse de la demande. Pour répondre à cette affluence, une API (Interface de programmation d'applications) a été créée pour faciliter l'accès aux bases Sirene. Tout utilisateur peut interroger la base et extraire les données souhaitées, le traitement étant automatisé et non plus à la demande. À l'avenir, Etalab, portail unique interministériel, pourrait regrouper l'ensemble des bases de données publiques y compris celles de Sirene.

Cet accès renforcé aux données ne concerne pas uniquement celles détenues par l'Insee, les entreprises sont également impactées par la loi pour une République numérique.

B - Une obligation de transmission de données étendue

La loi pour une République numérique modifie la loi du 7 juin 1951. Elle étend l'obligation des personnes morales de droit privé à transmettre des informations présentes dans leurs bases de données par voie électronique.

Concrètement, le Conseil national de l'Information Statistique (CNIS) autorise ces transmissions pour les besoins

d'enquêtes statistiques obligatoires. En cas de refus, les entreprises peuvent être mises en demeure et sont passibles d'une amende.

Ce changement de collecte concerne de nombreuses données parmi lesquelles on peut citer : les données de caisse, de sociétés d'énergie privées, de cartes bancaires, etc. Ce mode de transmission s'est accentué en période de crise sanitaire et a fortement contribué à éclairer la situation économique en France en 2020.

De plus, le processus est devenu plus simple et performant. Par exemple pour le calcul de l'IPC (Indice des Prix à la Consommation), 80 millions de produits sont suivis, contre 30000 auparavant quand ils étaient collectés par des enquêteurs pied.

Plus longs, moins coûteux, les traitements liés à ce nouveau mode de transmission s'en trouvent améliorés et modernisés.

II - Des traitements de données facilités

Que ce soit au sein de l'institut (A) pour la publication de ses indices, ou entre les différents acteurs du SSA^(B), certains traitements de données sont simplifiés grâce à la loi pour une République numérique.

A - De nouveaux traitements automatisés dans le calcul de l'IPC

Grâce à la transmission électronique des données de caisse, le calcul de l'indice des prix à la consommation s'est perfectionné et modernisé.

9075

Note de synthèse (administratif)

1/2

Les produits suivis dans le panier de l'IPC sont identifiés dans les données de caisse, fournies par les distributeurs, par leur code-barres. Cela permet plus facilement de classer les produits grâce notamment à un dictionnaire des codes barres.

Ainsi, grâce à ce dictionnaire, il est possible d'automatiser l'identification des produits, leur classement, et diminue considérablement le traitement manuel et donc la charge des enquêteurs.

En outre, les relances commerciales sont repérées automatiquement et donc la hausse de prix correspondante.

Enfin, il est beaucoup plus facile de remplacer un produit du panier IPC s'il disparaît en cours d'année. Grâce au dictionnaire des codes barres, il est possible d'automatiser le choix du produit le plus proche et éventuellement d'ajuster son prix.

Tous ces traitements sont rendus possibles par la loi pour une République numérique mais de nombreuses autres ont aussi été facilités au sein même du service statistique public.

B - Des appariements de données personnelles simplifiés dans le SSP

La loi du 7 octobre 2016 modifie celle du 6 janvier 1978 portant sur l'informatique, les fichiers et les libertés. Cela a pour but de permettre des appariements plus simples entre les différentes bases de données personnelles du SSP.

En effet, sur simple déclaration à la Cnil (Commission nationale de l'Informatique et des libertés), il est possible de traiter les données personnelles à des fins statistiques selon quelques conditions.

La première concerne l'identifiant des individus au RNIPP (répertoire national d'identification des personnes physiques) qu'on appelle le NIR. Ce NIR doit être remplacé par un code statistique non significatif (CSNS) commun à toute la statistique publique.

La seconde est que ce CSNS doit être cryptographié par l'Insee et que l'opération doit être renouvelée à une fréquence fixée par la Cnil.

Cette opération permet de nombreux appariements de sources du SSP, grâce à des fichiers pivot, et cela sécurise la diffusion des données.

Cette procédure de simple déclaration à la Cnil est également ouverte à la recherche scientifique et historique, alors qu'il était auparavant nécessaire pour eux et le SSP d'avoir une autorisation par décret d'Etat. Cependant, les scientifiques et historiens ne peuvent pas apparier leurs bases car le code crypté est unique à chaque projet.

Pour conclure, la loi pour une République numérique a eu un fort impact sur l'accès aux données, leur diffusion et leurs traitements pour le public comme les membres du service statistique public.

Cependant, dans une société de plus en plus numérique, il est certainement possible d'améliorer encore le processus.

9075

Note de synthèse (administratif)

2/2

Il serait dans ce ~~cadre~~ intéressant d'étudier les innovations récentes des services statistiques européens pour poursuivre les efforts entrepris dans le domaine de la numérisation.