

La révolution des données

Il existe de nombreuses définitions de la révolution des données. Le rapport du Groupe consultatif d'experts indépendants du Secrétaire général des Nations Unies (GCEU) évoque une « explosion » du volume et de la production des données associée à une « demande croissante de données dans l'ensemble de la société » (GCEU, 2014). L'approche complémentaire de PARIS21 est quelque peu différente, considérant qu'il s'agit de « fournir les bonnes données aux bonnes personnes au moment opportun » (PARIS21, 2015). Cette définition souligne le fait que la révolution des données devrait s'articuler autour d'une utilisation et d'un impact accrus des données.

A cet effet, dans la mesure du possible, les stratégies des systèmes de statistique nationaux doivent prévoir de nouvelles sources de données et un engagement accru auprès de nouveaux acteurs, tels que le secteur privé, les associations à but non lucratif et les universités. Les présentes lignes directrices mettent l'accent sur cet aspect particulier et primordial de la révolution des données. L'accès et l'utilisation de ces nouvelles sources de données dans un nouvel écosystème composé d'utilisateurs, de propriétaires et de producteurs de données, mais aussi de législateurs, qui permettront aux décideurs, aux organisations de la société civile et aux citoyens d'« effectuer un suivi des progrès en matière de développement, de responsabiliser les gouvernements et de favoriser le développement durable » (GCEU, 2014).

La révolution des données ne revêt pas la même signification selon la place occupée dans cet écosystème. Les systèmes de statistique officiels et nationaux doivent relever certains défis pour s'adapter à ce nouvel environnement de données. Les modèles de développement de la statistique mis en œuvre au cours des 15 à 20 dernières années peuvent être détournés par de nouveaux producteurs de données et risquent de devenir caducs si les pays ne s'adaptent pas. La Révolution des données affecte tous les domaines du système de statistique national. Dans certains pays, comme le Sénégal, des approches novatrices de la planification et l'adaptation des opérations statistiques fondées sur les registres d'appels téléphoniques des opérateurs de téléphonie mobile sous-tendent déjà de nouvelles pratiques. Dans d'autres pays aussi, les offices nationaux de statistique (ONS) devront s'adapter au nouvel écosystème pour rester à la page.

L'utilisation de nouvelles sources de données (cf. définition donnée ci-après dans cette section) est explicitement encouragée dans les Principes fondamentaux de la statistique officielle. En vue notamment de rendre effectif le droit d'accès des citoyens à l'information publique (suivant des critères de qualité, d'actualité et de coût), les ONS peuvent s'appuyer sur « toutes sortes de sources » (Nations Unies, 2014). Par définition, les caractéristiques officielles sont fournies par des instances officielles et suivant les normes et les standards professionnels décrits dans les Principes fondamentaux.

« Principe 5 : rapport coût-efficacité : les données utilisées à des fins statistiques peuvent être tirées de toutes sortes de sources [...]. Les organismes responsables de la statistique doivent choisir leur source en tenant compte de la qualité des données qu'elle peut fournir, de leur actualité, des coûts et de la charge qui pèse sur les répondants. » -- Nations Unies (2014), Principes fondamentaux de la statistique officielle.

Toutes les [phases de la SNDS](#) doivent répondre à ces nouvelles attentes en tenant compte des points suivants:

- Continuer à développer des systèmes de données administratifs permettant de réaliser des estimations de population fiables et solides. L'objectif est de rectifier les données disponibles sur la population et d'ancrer de nouvelles sources de données.
- Compléter les méthodes de collecte traditionnelles par de nouvelles sources de données sur la base des audits effectués. Il s'agit également de mettre en œuvre de nouveaux processus durables facilitant la collecte et favorisant la qualité des données obtenues.
- Entamer un processus d'évaluation adapté en vue de la substitution rentable des sources de données existantes.
- Élaborer un plan de données exhaustif, ainsi qu'une approche coordonnée de la collecte de données améliorant la fréquence des prévisions, optimisant la ventilation et la pertinence géographique (cf. feuilles de route pour la révolution des données élaborées par le Partenariat mondial pour les données de développement durable).
- Budgéter le personnel/les ressources humaines dans le domaine de la science des données, mais aussi les capacités juridiques et réglementaires.
- Élaborer un plan axé sur la conclusion de nouveaux partenariats, qui passe par des liens avec différents acteurs du secteur privé, par l'activation d'un réseau de consultants fiables ou d'agences régionales de statistique, l'objectif étant de faciliter l'accès aux grandes organisations multilatérales.
- Établir des liens forts avec les Stratégies régionales de développement de la statistique (SRDS) afin de combiner les ressources régionales au sein de pôles de compétences et d'excellence lorsque les agences de statistique nationales n'ont pas la capacité ni les ressources nécessaires pour s'adapter.
- Contrôler les processus de production statistique existants et les réviser si nécessaire

Améliorer les processus de données existants

Les données d'enquête et de recensement resteront sans doute la principale source d'informations pour le suivi international et les décisions nationales. Les systèmes de collecte de données administratives devront en outre être développés. Ces systèmes historiques perdureront et resteront pertinents pour rectifier les données disponibles sur la population. Des estimations de population fiables et solides permettront d'ancrer de nouvelles sources de données et d'en augmenter l'utilité. La révolution des données et les technologies connexes nous fournissent les outils permettant d'améliorer les mécanismes actuels de gestion des données:

1. **Méthodologie d'enquête.** Amélioration de la méthodologie d'enquête
2. **Collecte des données mobiles.** Saisie des données à distance sur des téléphones mobiles et des tablettes
3. **Données administratives.** Récente évolution de la disponibilité et de l'utilisation des données administratives (voir notamment la section Données ouvertes)
4. **Diffusion des données.** Un système de diffusion cohérent utilisant des outils innovants, basé sur les nouvelles technologies et évoluant en fonction de la

demande (lire notamment la section sur la diffusion de données)

L'introduction d'innovations peut contribuer à rationaliser les processus existants et à améliorer la fiabilité des données d'enquête. Les ONS doivent collaborer avec les différents secteurs et préparer, promouvoir et défendre le développement de systèmes administratifs forts. Dans de nombreux cas, les données ainsi obtenues seront moins chères à utiliser. La préparation de ces données réclame des efforts supplémentaires, même si elles ne sont pas basées sur la population. A cet effet, les ONS doivent recueillir et centraliser les listes d'infrastructures clés afin de contribuer à l'intégration du processus de planification. Les ONS doivent adopter un objectif clair : développer une zone de référence géospatiale avec des zones de recensement, des services de santé (cliniques et dispensaires), des écoles (primaires et secondaires) et des services de développement agricole délimités. L'objectif est d'obtenir une base de données de référence solide pour la planification et la stratification des enquêtes, mais aussi de fournir un service à la société civile.

Nouvelles sources de données et systèmes de statistique nationaux

Du côté des SSN, mobiliser la révolution des données en faveur du développement durable passe par l'exploitation d'informations en croissance exponentielle, majoritairement détenues par le secteur privé. De nouveaux partenariats conclus avec des organisations du secteur privé à but lucratif et non lucratif peuvent y contribuer, en aidant les SSN à faire des économies et à offrir des données plus détaillées et précises. Ils représentent toutefois aussi des risques et des défis (cf. PARIS21, 2015). Il s'agira de gérer des « big data » (également appelées mégadonnées ou données massives) -- « traces d'activités humaines relevées par des dispositifs numériques » (Letouzé et al., 2013) -- et probablement de nouer des partenariats indispensables entre les universités, les analystes politiques et les SSN. Il conviendra de procéder avec minutie, dans la mesure où l'utilisation de ces données requiert l'emploi de techniques analytiques relativement sophistiquées.

De nouvelles données seront probablement définies. Cette section privilégie une approche plus pragmatique qu'une tentative de description des éléments constitutifs d'une nouvelle source de données. Le champ des possibles est ainsi réduit aux cinq sources de statistiques officielles généralement qualifiées de « nouvelles », présentées par ordre de faisabilité dans l'optique d'une mise en œuvre dans un pays en développement.

1. **Données de capteurs et données géospatiales**. Exemple : utilisation de photos prises par satellite pour estimer les niveaux de pauvreté (cliquer [ici](#) pour en savoir plus).
2. **Données de télécommunications**. Exemple : utilisation du détail des appels pour évaluer le niveau de pauvreté et de santé (cliquer [ici](#) pour en savoir plus).
3. **Transactions commerciales, y compris données obtenues par lecture optique, données de carte de crédit, et c.** Exemple : utilisation de données obtenues par lecture optique pour le calcul des indices de prix à la consommation (cliquer [ici](#) pour en savoir plus).
4. **Web crawling, scraping, recherche et analyse**. Exemple : utilisation des offres d'emploi mises en ligne pour effectuer une estimation du taux de chômage ou les données LinkedIn pour prévoir les mutations au sein de catégories professionnelles (cliquer [ici](#) pour en savoir plus).
5. **Réseaux sociaux**. Exemple : utilisation des tendances Google et d'analyses de sentiments pour mesurer le bien-être subjectif (cliquer [ici](#) pour en savoir plus).

Ces sources de données sont particulièrement utiles pour renseigner les indicateurs entre deux enquêtes et pour détecter les changements intervenus au niveau des indicateurs à évolution rapide. L'étude du cas de certains pays offrira une source d'informations essentielle, l'utilisation de ces données étant extrêmement limitée à l'heure actuelle. Le SSN a un rôle à jouer dans le développement d'une meilleure compréhension des applications nationales. Une participation active des institutions régionales sera probablement requise pour gérer les ressources rares et exploiter des économies d'échelle.

Les modalités d'accès aux données massives (détenues par le secteur privé) et les problèmes de confidentialité associés diffèrent de ceux liés à l'utilisation des données administratives (également parfois qualifiées de « massives »). Pour accéder aux données administratives, les ONS peuvent s'appuyer sur les cadres juridiques existantes. Les données d'entreprise, en revanche, sont un terrain vierge. Les modalités d'accès à ces données devront être définies avec les conseils nationaux sur la protection de la vie privée et toutes les parties prenantes. Au niveau international cependant, le Groupe de travail mondial des Nations Unies sur l'utilisation des mégadonnées en statistique officielle travaille actuellement à l'élaboration de principes pour l'accès à ces données. Ce travail constituera un complément utile aux Principes fondamentaux de la statistique officielle (Nations Unies, 2014).

Robin, Klein et Jütting (2016) fournissent une présentation détaillée des avantages et des complémentarités, ainsi que des risques et des défis associés à l'utilisation de nouvelles sources de données en statistique officielle. Les points suivants résument les principales leçons à tirer pour les SSN.

Avantages et complémentarités

- **Rapport coût-efficacité.** Les partenariats public-privé (accords collaboratifs volontaires visant à accroître la capacité des SSN à fournir des statistiques améliorées ou inédites) peuvent aider les SSN à économiser des ressources, grâce au partage de données et à l'absence de coûts d'investissement en infrastructures de gestion des données. Les coûts marginaux du transfert des données collectées par le secteur privé à une partie prenante du SSN sont très faibles. Alors qu'une enquête coûte plus de 20 millions de dollars aux Etats-Unis, la mise en correspondance de microdonnées privées avec des données agrégées existantes (par ex. informations relatives à une usine et données de l'entreprise) ne mobiliserait qu'un cinquième de ce montant (Landfeld, 2014). Par ailleurs, en externalisant le traitement des données, une partie prenante de SSN aux ressources limitées peut tirer parti des logiciels et de l'expertise du secteur privé et éviter des coûts d'investissement élevés.
- **Actualité.** Les métadonnées mobiles brutes étant disponibles de façon quasi-instantanée, les registres d'appels des opérateurs de téléphonie mobile, par exemple, peuvent générer des statistiques en temps quasi-réel.
- **Niveau de détail.** Les données du secteur privé – en particulier les registres d'appels et les données géospatiales – offrent un niveau de précision temporelle, spatiale, thématique et individuelle élevé, qui peut être utile pour l'évaluation des politiques à court terme et la production de statistiques ventilées aux niveaux régional et infranational, par exemple.

De nouvelles sources de données permettent aux agences de statistique d'établir des tendances auparavant considérées comme non mesurables et d'améliorer leur réactivité en

cas de brusque changement de politique.

- **Données relatives à de nouveaux domaines.** Les données massives (big data) revêtent un potentiel en termes d'aide à la création de nouveaux indicateurs, auparavant non calculés par les ONS, notamment pour la mesure des inégalités pertinentes dans le cadre des ODD.
- **Réactivité accrue.** De nouvelles sources de données doteront les ONS des capacités requises pour réagir rapidement à de nouvelles conditions et aider les universités à réaliser des projections.

Risques et défis

Quatre défis, liés aux propriétés des données concernées, distinguent la plupart des partenariats du domaine de la statistique de ceux conclus dans d'autres secteurs comme la santé ou les infrastructures : la garantie de la sécurité des données protégées, la création d'un modèle commercial de partage des données, la protection de la vie privée et la gestion des difficultés techniques associées aux mégadonnées.

- **Accès.** La divulgation d'informations protégées est perçue comme un risque majeur pour les organisations à but non lucratif comme à but lucratif. Les données qui contiennent des informations concrètes sur les clients ou la stratégie d'une organisation sont souvent confidentielles. Les registres d'appels utilisés par les entreprises à des fins de géo-marketing sont ainsi beaucoup plus sensibles que les tweets publics, relativement accessibles. D'autres préoccupations concernent la possibilité que les gouvernements utilisent les informations à des fins réglementaires ou que la publication des données relatives aux clients d'une organisation puisse nuire à leur image.
- **Incitations et durabilité.** Certains facteurs peuvent nuire à l'attrait du modèle de partenariat en matière de données. L'incertitude relative à la demande de données massives peut ainsi être source de doutes quant à leur portée. De plus, les avantages du partenariat ne sont pas toujours immédiats ou directs. Enfin, la durabilité des nouvelles sources de données est incertaine. Compte tenu du fait que les données privées sont initialement collectées à des fins non statistiques, la gestion du processus d'extraction de ces données peut en effet être lourde si le champ d'application initial perd de son intérêt.
- **Vie privée et éthique.** La dimension de partage des données propre au concept de partenariat est une menace pour la vie privée des individus ou de catégories d'individus. Par conséquent, la protection des informations à caractère personnel ou relatives à un groupe d'individus est donc à la fois une condition et un objectif proprement dit de la mise en œuvre des partenariats en matière de données. Les législations sur la protection de la vie privée sont souvent sources de contraintes réglementaires. La plupart des législations actuelles applicables en la matière ne couvrent pas spécifiquement les données massives. Les lois existantes offrent une marge d'interprétation. Les ONS n'ont pas clairement pour vocation d'exploiter les microdonnées sensibles telles que les registres d'appels. De plus, les acteurs publics et privés sont confrontés à des problèmes éthiques et de réputation : le simple fait qu'une entreprise conserve les données relatives à ses clients peut pousser ces derniers à changer de fournisseur. Le transfert de ces données représente donc un risque majeur pour les organisations.
- **Défis techniques et statistiques.** Ces défis sont liés à la nature même de la

plupart des données à caractère privé, particulièrement des mégadonnées, qui nécessitent souvent des infrastructures spécialisées et ne sont pas toujours centralisées, standard, structurées et représentatives. Les caractéristiques des jeux de données massives imposent donc des restrictions au niveau des caractéristiques structurelles des partenariats conclus, mais aussi du type de statistiques qui peut être produit.

Intégration de nouvelles sources de données dans une SNDS

Le rapport du Groupe d'experts des Nations Unies et de l'extérieur sur la révolution des données prône explicitement une modification de l'approche des SNDS tenant compte de la révolution des données. Il s'agit de

[...] transposer à plus grande échelle les stratégies nationales de développement de la statistique (SNDS) pour améliorer la planification à long terme et sa coordination, d'identifier les bons investissements et d'associer les producteurs de données non officiels à un travail de coopération visant à accélérer la production, la diffusion et l'utilisation de données, mais aussi de consolider les capacités et les ressources de la société civile affectées à la production, l'utilisation et la diffusion des données. – GCEU (2014, page 25)

La révolution des données transformera la façon dont les ONS et les SSN opèrent et nécessite l'implication de nouveaux acteurs dans le processus d'élaboration de la SNDS.

- **Évolution du rôle des ONS.** La transformation de l'écosystème des nouveaux producteurs et utilisateurs de données débouchera sur l'évolution des modèles d'activité des ONS et sur l'émergence d'autres producteurs. Les ONS seront moins intégrées verticalement et externaliseront davantage leurs processus statistiques. Cela va de pair avec une évolution du rôle de responsable de la production statistique des ONS à celui d'organe chargé de la gestion des défis liés à l'évaluation des risques et des coûts.
- **Évolution des profils de compétence.** L'évolution du rôle des ONS s'accompagne de la création de nouvelles exigences relatives à l'éventail des compétences dont ils doivent disposer. Le personnel des ONS doit être correctement formé aux nouvelles méthodologies pour pouvoir identifier, évaluer et utiliser les nouvelles sources de données. Cela implique des compétences et des capacités de formation dans le domaine naissant de la science des données, mais aussi des capacités juridiques et réglementaires.
- **Création de centres régionaux d'assistance.** Là où les organes de statistique nationaux n'ont pas les capacités et les ressources nécessaires pour s'adapter, la SRDS servira de référence pour l'approche coordonnée de la transposition de la révolution des données dans la statistique nationale. Domaines dans lesquels la révolution des données pourrait être mise à profit pour faire progresser le changement au niveau régional : (i) création de pôles d'excellence et de compétence, , (ii) mise à disposition de « bacs à sable » contenant des ensembles de mégadonnées: utilisation de plates-formes évolutives telles que celles de la [UNECE](#) pour explorer les jeux de données étendus d'une organisation de façon interactive et collaborative, (iii) concentration des ressources en vue de la conclusion de partenariats universitaires clés et de la promotion de partenariats public-privé, afin de contribuer au pool d'expertise régional.
- **Une approche mixte de la compilation de statistiques officielles** . la capacité

de généralisation statistique de nombreuses sources de données non traditionnelles n'est pas encore très bien maîtrisée. Il convient donc de les utiliser avec précaution et de recourir à des sources classiques pour valider et étalonner les estimations effectuées, notamment à court terme. Une telle approche mixte et complémentaire implique que les ONS continuent à s'appuyer sur des méthodes statistiques traditionnelles.

- **Protocoles.** Le succès des partenariats dépend de l'adoption d'approches systématiques et transparentes d'un partage de données fondé sur des protocoles. Ces approches doivent permettre de limiter les risques de ré-identification des individus concernés. De tels protocoles sont déjà en place pour les données médicales sensibles. Ils sont essentiels pour instaurer la confiance dans la fiabilité et l'intégrité des systèmes de statistique nationaux s'agissant des données communiquées involontairement.
- **Montrer l'exemple.** Les divers acteurs du SSN utilisent différentes sources à des rythmes variables. L'ONS est souvent l'organe responsable de la formulation et de la mise en œuvre de la SNDS du pays en question. Les ONS ont un rôle d'exemple important à jouer quant à la façon d'utiliser les nouvelles sources de données en statistique officielle. Ils doivent s'efforcer d'expérimenter de nouvelles sources tout en prêtant l'attention qui convient aux problématiques liées à la confidentialité et à la qualité des données.

De nouvelles formes de partenariat

L'accès à de nouvelles sources de données implique de nouvelles formes de partenariat. Les dernières années ont vu l'émergence de plusieurs structures coopératives fructueuses, qui réunissent généralement des acteurs du secteur privé. Trois caractéristiques spécifiques distinguent ces partenariats public-privé (PPP) propres au domaine de la statistique de ceux conclus dans d'autres secteurs :

1. Ils doivent donner lieu à des contrats conclus à long terme, dans la mesure où les besoins concernent généralement des données longitudinales et où il existe peu de fournisseurs concurrents (les registres d'appels téléphoniques, par exemple, sont souvent détenus par un petit nombre d'opérateurs de réseaux mobiles).
2. Les risques liés à la confidentialité et aux informations protégées jouent un rôle essentiel dans les PPP conclus dans le domaine de la statistique, alors que dans d'autres secteurs (infrastructures notamment), les risques sont principalement limités au rapport qualité-prix et au retour sur investissement.
3. Les PPP conclus dans le domaine de la statistique couvrent toutes les phases de la chaîne de création de valeur (collecte, traitement, analyse et diffusion des données).

Ces partenariats sont parfois longs à établir. Par conséquent, les ONS doivent tirer le meilleur parti des structures en place, par exemple en puisant dans un réseau de tiers ou en explorant des sources de données moins sensibles. Il y a également une carte importante à jouer en faveur d'une coopération plus étroite entre les ONS et les instances de statistique régionales. Ces dernières ouvrent souvent la voie à de vastes coopérations multilatérales et réduisent les coûts de coordination. Il y a aussi une marge de renforcement de la collaboration entre les ONS des pays en développement et des pays développés, qui peut passer par le partage de photos prises par satellite.

Cadres légaux

Tout obstacle à l'obtention et à l'utilisation de nouvelles sources de données correspond à un besoin de créer des instruments et des procédures juridiques permettant d'accéder aux données et de préserver la confidentialité des utilisateurs. Ces utilisateurs sont du domaine privé et forment une source très abondante de données clients protégées par des intérêts privés. Dans de nombreux pays, la loi ne définit pas clairement les procédures juridiques visant à obtenir ces données et à faire en sorte qu'elles relèvent du domaine public. La période des OMD a permis une prise de conscience, avec la mise à disposition en ligne de microdonnées d'enquêtes pour les chercheurs. Cela a impliqué la définition d'un protocole de documentation des données et de normes. Les définitions de métadonnées ont ainsi évolué (voir projet DDI et initiative SDMX). De plus, de nombreux travaux de plaidoyer ont été entrepris pour modifier les structures juridiques et redéfinir les droits des utilisateurs régissant les processus de garantie de la confidentialité. Les instruments tels que la Charte africaine de la statistique ont fourni un cadre légal aux pays pour leurs processus législatifs. La charte contient en effet des lignes directrices servant de guide pour la modification de la législation régissant la production de nouveaux types de données, comme les données massives (ou big data).

Outils:

La liste ci-dessous répertorie les outils de gestion des données et les nouvelles sources de données disponibles:

L' [inventaire des projets relatifs aux mégadonnées](#) réalisé par le Groupe de travail mondial des Nations Unies dresse un catalogue de tous les projets pertinents pour la statistique officielle. Il recense également les indicateurs relatifs aux ODD et d'autres statistiques requises pour prendre des décisions éclairées en matière de politiques publiques, ainsi que pour la gestion et le suivi des programmes / projets du secteur public.

[ADAPT](#) (acronyme de l'anglais Advanced Data Planning Tool) est un outil de planification innovant destiné aux offices de statistique désireux de s'adapter à l'évolution de la demande et des pratiques en matière de données. Cet outil permet aux producteurs de données du système de statistique national de consulter, budgéter et visualiser de façon graphique les indicateurs définis par le plan national de développement. Il est destiné aux pays qui cherchent à remplir les exigences fixées par les instances internationales en matière de suivi des ODD et à les aborder au regard de leurs propres priorités nationales.

La [Plate-forme pour les innovation en matière de statistique](#) (PISTA) répertorie une série d'innovations proposées dans le domaine des données et de la statistique officielle dans les pays en développement. Il fournit des informations de base, telles que des comptes rendus, des coordonnées, de brèves appréciations et des études de cas sur des innovations institutionnelles, organisationnelles et technologiques des secteurs public et privé.

L'initiative [Global Innovations Exchange de l'USAID](#), l'agence américaine pour le développement international, constitue une plate-forme mondiale en ligne dédiée à l'innovation, au financement et aux ressources en matière de développement. Elle a pour but de permettre aux innovateurs d'accéder aux ressources, aux contacts et aux informations nécessaires pour mener leurs projets à maturité. Le [GPSDD Toolbox](#) (Partenariat mondial pour les données du développement durable) regroupe un ensemble d'outils, de méthodes et de ressources. Ce module a été mis au point par des acteurs de terrain issus de toutes les régions du monde et utilisant des données sur le développement.

Bonnes pratiques:

Un certain nombre de projets collaboratifs de partage de données entre le secteur privé et le secteur public ont d'ores et déjà été entamés. Selon Robin, Klein et Jütting (2016), on

peut ainsi classer les projets dans plusieurs catégories génériques :

[Production de statistiques en interne](#) : l'opérateur de téléphonie mobile Telefónica a utilisé ses registres d'appels pour développer plusieurs applications en recourant à ses capacités internes. Ces projets démontrent que les producteurs de données privés sont non seulement désireux de contribuer à la réduction des lacunes en matière de statistique, mais peuvent aussi bénéficier de l'utilisation de leurs données et de leurs ressources dans l'intérêt du public.

[Transfert des jeux de données aux utilisateurs](#) : selon ce modèle, les informations sont transmises directement du propriétaire des données à l'utilisateur final. Ce dernier bénéficie ainsi d'une plus grande flexibilité d'utilisation des données. Les données brutes sont anonymisées, échantillonnées et ventilées pour éviter toute possibilité de ré-identification. De 2012 à 2015, l'opérateur de téléphonie mobile Orange a organisé deux défis d'innovation. Dans ce cadre, Orange a mis ses registres d'appels à disposition des équipes de recherche du monde entier, sous une forme anonymisée, en dépit des risques liés à la confidentialité et aux informations protégées.

Accès à distance : selon ce modèle, les propriétaires des données offrent aux utilisateurs finaux un accès complet à celles-ci tout en assurant un contrôle très strict des informations extraites des bases et des jeux de données. Il existe plusieurs exemples de systèmes d'accès à distance, à l'instar de l'initiative [Data for Good](#) de la société Real Impact Analytics. Dans le cadre de ce projet, l'entreprise accède aux données de télécommunication contenues dans l'environnement sécurisé des opérateurs.

[Transfert des jeux de données à un tiers de confiance](#) : ni le propriétaire, ni l'utilisateur des données ne portent ici le fardeau de l'hébergement sécurisé des données. Les deux parties s'en remettent à un tiers de confiance pour assurer cette tâche et les services nécessaires pour permettre un accès sécurisé à la source de données. Depuis 2009, des statistiques de déplacements basées sur des registres d'appels sont utilisées pour calculer le solde de comptes de voyage, grâce à un partenariat public-privé conclu entre la société Positium et la Banque centrale d'Estonie (EestiPank).

Transfert d'algorithmes et non de données : le modèle des algorithmes partagés permet à plusieurs propriétaires de données privés souhaitant exécuter des fonctions analytiques similaires de réutiliser un logiciel. Le projet [Open Algorithms](#) (OPAL) vise ainsi à exploiter le potentiel des données privées en mettant à disposition une plate-forme en accès libre et des algorithmes prêts à l'emploi. Des entreprises privées peuvent ainsi exécuter des algorithmes prédéfinis en toute autonomie dans leurs propres environnements sécurisés et publier uniquement les résultats agrégés.

Références:

GCEU (2014). [Un monde qui compte : Mobiliser la révolution des données pour le développement durable](#), Groupe consultatif d'experts indépendants sur la révolution des données pour le développement durable.

Landfeld, S. (2014). [Uses of Big Data for Official Statistics: Privacy, Incentives, Statistical Challenges, and Other Issues](#). Dans : United Nations Statistics Division (UNSD) and National Bureau of Statistics of China, International Conference on Big Data for Official Statistics, Beijing, China: 28-30 octobre 2014

Letouzé et al. (2013), [Big Data for Conflict Prevention: New Oil and Old Fires](#). Dans : Francesco Mancini, ed., New Technology and the Prevention of Violence and Conflict, New York: International Peace Institute, avril 2013.

La révolution des données

NSDS GUIDELINES (<https://nsdsguidelines.paris21.org>)

PARIS21 (2015). [Feuille de route pour une révolution des données menée par les pays](#), secrétariat de PARIS21.

Robin, N., T. Klein et J. Jutting (2016). [Public-Private Partnerships for Statistics. Lessons Learned, Future Steps: A focus on the use of non-official data sources for national statistics and public policy](#), OECD Development Co-operation Working Papers, No. 27, OECD Publishing, Paris. <http://dx.doi.org/10.1787/5jm3nqp1g8wf-en>

Nations Unies (2014), [Principes fondamentaux de la statistique officielle](#), Résolution 68/261 adoptée par l'Assemblée générale, Nations Unies, <http://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf>.

Source URL: <https://nsdsguidelines.paris21.org/fr/node/716>