



00

Bases statistiques et généralités

1788-1700

Office fédéral de la statistique Stratégie d'innovation sur les données

Buts, objectifs stratégiques,
étapes de réalisation



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Département fédéral de l'intérieur DFI
Office fédéral de la statistique OFS

Neuchâtel 2017

Éditeur:	Office fédéral de la statistique (OFS)
Renseignements:	Bertrand Loison, OFS, tél. 058 463 67 70, bertrand.loison@bfs.admin.ch
Rédaction:	Bertrand Loison, OFS; Diego Kuonen, Statoo Consulting
Série:	Statistique de la Suisse
Domaine:	00 Bases statistiques et généralités
Langue du texte original:	anglais
Traduction:	Services linguistiques de l'OFS
Mise en page:	section DIAM, Prepress/Print
Page de titre:	OFS; concept: Netthoevel & Gaberthüel, Bienne; photo: © vinnstock – Fotolia.com
Copyright:	OFS, Neuchâtel 2017 La reproduction est autorisée, sauf à des fins commerciales, si la source est mentionnée.
Commandes d'imprimés:	Office fédéral de la statistique, CH-2010 Neuchâtel, tél. 058 463 60 60, fax 058 463 60 61, order@bfs.admin.ch
Prix:	gratuit
Téléchargement:	www.statistique.ch (gratuit)
Numéro OFS:	1788-1700

Table des matières

Note préliminaire	5
Introduction	7
Notre vision de l'innovation sur les données	9
Bases légales actuelles	9
Données primaires et données secondaires	10
Le cycle induction-déduction	11
Premier objectif stratégique	11
Deuxième objectif stratégique	13
Considérations générales	13
Réalisation des objectifs stratégiques	14

Note préliminaire

L'idée essentielle consiste à augmenter et/ou compléter la production statistique officielle de base de l'Office fédéral de la statistique (OFS) partout où l'innovation sur les données (définie ci-après) fait sens. La portée de cette stratégie est toutefois limitée. Elle ne prétend pas embrasser tous les aspects de l'innovation à l'OFS et elle ne vise pas à établir un ordre de priorités. La stratégie sera progressivement révisée et mise à jour au fur et à mesure de la réalisation des objectifs liés aux étapes définies ci-après.

Introduction

La révolution numérique, qui est loin d'être terminée, s'appuie sur des données: des termes tels que *big data*, *cloud*, *internet of things*, *internet of everything*, *quatrième révolution industrielle*, *smart cities* et *data economy*, qui sont sur toutes les lèvres, renvoient à des notions qui bouleversent nos habitudes de consommateurs et d'entrepreneurs. La transformation numérique a commencé avec la première vague de numérisation: la numérisation technique, qui a consisté à convertir des contenus et des services analogiques en des contenus et des services numériques (sans changements fondamentaux dans les processus sous-jacents), et qui a débouché sur la révolution du *big data*. Le *big data* est une infrastructure de gestion des données (et des matériels, logiciels et architectures sous-jacents). Elle n'est pas un «musée» voué à la seule conservation des données. Le statisticien officiel W. Edwards Deming notait en 1942 déjà que «le but ultime de toute collecte de données est de fournir des bases ou des recommandations en vue d'une action». Une deuxième vague de numérisation est nécessaire pour nous permettre de créer, à partir de la révolution du *big data*, des connaissances et de la valeur pour la société: c'est ce que nous appelons **innovation sur les données**. En ce domaine, le principal défi à relever est celui de la véracité des données, fondement de la **confiance dans les données**, qui comprend la fiabilité, la solidité et la validité des données, leur qualité, et la transparence des processus de production des données. La **vérité des données** est plus importante que jamais dans notre monde de «post-vérité» et de fausse information.

La statistique publique a toujours utilisé une grande diversité de sources de données. Depuis sa création en 1860, l'OFS s'attache à évaluer constamment la véracité et le «pédigrée» de ses données pour pouvoir fournir des services statistiques fiables et dignes de confiance. Mais l'utilisation (ou la réutilisation) des sources de données existantes et/ou l'utilisation de sources de données nouvelles pourrait fournir à l'OFS une opportunité d'innover, d'augmenter et/ou de compléter ses statistiques publiques, de fournir un service statistique plus efficace, plus adapté aux besoins du moment, plus efficient et plus complet.

Il importe de souligner que la révolution des données ne se limite pas au monde de la statistique (officielle).

Au niveau international, la possibilité pour la statistique d'accéder à des sources de données nouvelles fait l'objet de discussions dans les conférences statistiques globales et régionales. L'«Agenda 2030 pour le développement durable et les objectifs du développement

«durable» ont fait évoluer les choses: les données et les statistiques (officielles) retiennent désormais beaucoup plus l'attention du monde politique. En raison de ces demandes, les possibilités d'accès à de nouvelles sources de données et l'utilisation des données nouvelles doivent être discutées, testées et si possible mises en pratique. Le «*Cape Town Global Action Plan for Sustainable Development Data*», adopté dans le cadre de la 48^e session de la Commission statistique de l'ONU en mars 2017, appelle différentes actions en relation avec les nouvelles sources de données. Il est nécessaire, par exemple, d'identifier et de lever les barrières qui s'opposent à l'utilisation des nouvelles sources de données – registres officiels, données administratives et autres données issues de sources nouvelles – et il faut coordonner les efforts visant à les incorporer dans la production statistique officielle. Le travail de l'OFS y contribue.

Au niveau national, différentes stratégies visant à promouvoir l'utilisation de nouvelles sources de données et des registres officiels ont été formulées ces dernières années (p. ex. «Stratégie Suisse numérique», «politique nationale des données»). Toutefois, ces stratégies ont généralement été élaborées dans une optique technologique et elles mettent peu l'accent sur l'utilisation des données à des fins statistiques. Comme la statistique officielle et en particulier l'OFS subissent une pression croissante du Parlement visant à réduire la charge que la statistique fait peser sur les personnes interrogées – par l'utilisation des registres officiels, des données administratives et d'autres sources potentielles de données résultant des progrès de la numérisation –, le travail de l'OFS consistera à la fois à aider à positionner la statistique officielle dans la mise en œuvre de ces initiatives stratégiques et à contribuer au développement de solutions efficaces pour répondre aux pressions et aux attentes du monde politique.

Notre vision de l'innovation sur les données

Si l'on entend par **analyse** l'acquisition de connaissances (ou l'interprétation d'un phénomène) à partir d'un ensemble de données, on peut définir

l'**innovation sur les données** comme l'application de *méthodes d'analyse complémentaires* (p. ex. l'analyse prédictive par des techniques statistiques avancées, la science des données et/ou l'apprentissage automatique) aux données provenant de sources existantes (ou traditionnelles) et/ou à des données issues de sources nouvelles (non traditionnelles)

pour aider la statistique publique à jouer son rôle dans le processus démocratique en garantissant que l'information produite demeure fiable, transparente et digne de confiance.

L'innovation sur les données pourrait offrir des opportunités pour des interventions plus avantageuses en termes de rapport coût/bénéfice et elle pourrait mettre en évidence des faits omis par les approches et les méthodes d'analyse traditionnelles. L'innovation sur les données vise à augmenter la qualité, la portée et le rapport coût/efficacité des produits statistiques tout en réduisant la charge imposée aux ménages et aux entreprises interrogés. L'innovation sur les données comporte aussi un potentiel d'amélioration de la pertinence et du degré de détail des statistiques publiques.

Bases légales actuelles

Selon la loi sur la statistique fédérale, l'OFS et les autres services de statistique publique ont pour mission de produire de manière professionnelle et en toute indépendance des informations représentatives sur l'état et l'évolution de la population, de l'économie, de la société, de la formation, de la recherche, du territoire et de l'environnement en Suisse. Pour remplir cette mission, l'OFS utilise plusieurs sources de données différentes. En vertu de la loi, l'OFS doit vérifier si la Confédération, les cantons, les communes ou d'autres entités de droit public possèdent des registres ou des données administratives de qualité suffisante pour pouvoir être utilisées à des fins statistiques. Si ce n'est pas le cas, et seulement pour les informations statistiques produites en vertu du programme pluriannuel, l'OFS collecte des données par voie d'enquêtes ou de recensements. Le nombre et le type des enquêtes sont limités au strict nécessaire. Les données collectées par l'OFS présentent l'avantage d'obéir à un projet statistique clair et d'être sous le contrôle de l'OFS, qui garantit la véracité et la qualité des données.

Données primaires et données secondaires

Dans le contexte de l'innovation sur les données, l'OFS divise ses sources de données en données primaires et données secondaires.

Les données dites **primaires** sont collectées (et planifiées) par l'OFS à des fins statistiques pour expliquer et vérifier la validité d'une hypothèse, par l'opérationnalisation d'une théorie. L'acquisition de connaissances à partir de données primaires est ce qu'on appelle l'analyse primaire («top-down», c'est-à-dire explicative et confirmatoire). L'analyse procède «par déduction» à partir d'une hypothèse (ou d'une théorie) initiale. Exemples de données primaires: les données issues de sources traditionnelles telles que les recensements et les enquêtes statistiques, qui ont été collectées par l'OFS à des fins statistiques.

Les données dites **secondaires** sont des données qui ont été collectées (et planifiées) à d'autres fins (et qui souvent ne sont pas «sous le contrôle» de l'OFS) mais qui peuvent être utilisées pour produire des idées (ou des «théories») nouvelles. L'acquisition de connaissances à partir de données secondaires est ce qu'on appelle l'analyse secondaire («bottom-up», c'est-à-dire exploratoire et prédictive). Elle s'appuie sur le «raisonnement par induction», dans lequel les données sont premières. Exemples de données secondaires: les données issues de sources non traditionnelles telles que les registres internes ou externes de l'OFS, les données administratives et les données numériques provenant d'instruments de mesure, de machines, de senseurs, de satellites, de drones ou des médias sociaux. Les sources secondaires n'ayant pas été planifiées pour entrer directement dans le système de production de la statistique publique, elles doivent être traitées pour «permettre des inférences statistiques», autrement dit pour qu'on puisse en tirer des conclusions, par le raisonnement déductif, dans le cadre de la statistique publique.

Les données secondaires peuvent être classées en **données identifiables** – données associées à une unité singulière située dans l'espace ou dans le temps, p. ex. un individu, une institution, un produit ou un lieu géographique (exemples: données des registres officiels, données administratives, données de l'imagerie satellite, informations géospatiales, codes-barres de produits) – et en données **non identifiables**, qui ne sont pas identifiables à un niveau quelconque de leur utilisation (p. ex. les tendances Google, les contenus de Twitter et les données issues d'autres types de médias sociaux). Les données secondaires identifiables sont utilisables à des fins d'inférence statistique si leur véracité a pu être établie (ainsi les données des registres couramment utilisés par l'OFS). Les données secondaires non identifiables sont d'une utilité limitée pour l'inférence statistique car leur véracité ne peut pas être établie avec certitude.

Le cycle induction-déduction

Il importe de noter que les deux approches de l'analyse (raisonnement inductif et raisonnement déductif) sont *complémentaires et utilisées de manière itérative* aux fins d'améliorer constamment les décisions politiques et de rendre possibles des politiques fondées sur des données objectives. Les méthodes analytiques aujourd'hui utilisées par l'OFS continueront à être utilisées en parallèle avec des méthodes d'analyse complémentaires.

Par exemple, l'information produite par le raisonnement inductif ne peut pas expliquer si et pourquoi ces découvertes sont utiles et dans quelle mesure elles sont valides. Les outils de confirmation fournis par le raisonnement déductif permettront alors de confirmer les découvertes et d'évaluer la qualité des décisions fondées sur ces découvertes.

Premier objectif stratégique

Le premier objectif stratégique énoncé dans le programme pluriannuel 2016–2019 est d'assurer la production statistique de manière durable et d'adapter en permanence les statistiques aux besoins des utilisateurs. L'une des mesures citées pour atteindre cet objectif consiste à élaborer des lignes directrices sur l'utilisation potentielle de nouvelles sources de données.

En conséquence, le premier objectif de la présente stratégie d'innovation sur les données consiste à faire prendre conscience que l'innovation sur les données **est une question stratégique majeure**.

Ce premier objectif stratégique sera révisé progressivement et mis à jour au fur et à mesure de la réalisation des étapes définies plus bas.

Objectif stratégique 1: élaborer des lignes directrices en matière d'innovation sur les données et étudier la faisabilité de l'application de méthodes d'analyse complémentaires aux sources de données existantes (ou traditionnelles) et/ou à des sources de données nouvelles (non traditionnelles), afin d'augmenter et/ou compléter la production statistique de base existante dans les domaines où l'innovation sur les données est souhaitable.

La **1^{re} étape** consistera à appliquer, dans des projets pilotes internes à l'OFS, des méthodes d'analyse complémentaires aux sources de données primaires existantes (traditionnelles) de l'OFS et aux données déjà croisées de sources secondaires identifiables (s'il en existe). Le but est d'augmenter et/ou compléter ce qui existe, autrement dit d'utiliser les sources de données existantes de manière innovante afin d'acquérir de l'expérience pratique et pour dresser un inventaire des défis à relever ainsi que des ressources, des compétences et des technologies nécessaires pour mettre en œuvre l'innovation sur les données et explorer les possibilités de «gains immédiats».

Si la faisabilité du projet est démontrée, une **2^e étape** pourrait consister à compléter et/ou augmenter la production statistique de l'OFS par des données produites par les méthodes d'analyse complémentaires appliquées à des données secondaires additionnelles, déjà utilisées à l'OFS.

Une **3^e étape** pourrait consister à appliquer des méthodes d'analyse complémentaires à des données secondaires nouvelles (non encore utilisées à l'OFS) pour investiguer et produire de nouvelles informations statistiques et de nouvelles statistiques dans des domaines statistiques particuliers.

Voici l'ordre dans lequel il serait préférable d'utiliser les sources de données dans le cadre de la stratégie d'innovation sur les données:

1. Sources primaires de l'OFS et sources secondaires identifiables déjà croisées (si elles sont déjà utilisées dans la production statistique actuelle de l'OFS);
2. Données secondaires additionnelles déjà utilisées à l'OFS;
3. Sources secondaires nouvelles (non utilisées jusqu'ici à l'OFS).

Cet ordre d'utilisation des sources vise à accroître l'efficacité de la production statistique, à limiter les coûts et à réduire la charge administrative imposée aux entreprises et aux individus. Cet ordre permettra d'acquérir une meilleure compréhension des méthodes, des technologies et des outils disponibles, sans se lancer dans des opérations trop complexes, comme ce serait le cas si l'on commençait directement par la 3^e étape, ce qui soulèverait à la fois des problèmes légaux, des difficultés technologiques (IT) et des problèmes de capacité. Enfin, se familiariser avec les méthodes d'analyse complémentaires, c'est aussi augmenter l'aptitude des statisticiens à rechercher de nouvelles sources de données adaptées aux objectifs spécifiques de la statistique publique.

Deuxième objectif stratégique

L'innovation sur les données suppose un changement de paradigme dans la mesure où elle consiste à associer le raisonnement déductif et le raisonnement inductif, et elle conduira nécessairement à un nouveau mode de production et de communication dans la statistique publique.

Il importe de souligner qu'il n'existe pas d'approche méthodologique générale applicable à toutes les formes de mise en œuvre des méthodes d'innovation sur les données présentées plus haut. Chaque domaine d'application et chaque type de données comporte ses propres défis et potentialités. Les processus et les modèles d'activité de l'OFS devront être adaptés et étendus pour intégrer l'innovation sur les données.

La **communication** sur toutes ces questions jouera un rôle déterminant. D'où le deuxième objectif stratégique.

Objectif statistique 2: développer et implémenter des mesures de communication internes et externes sur la valeur ajoutée que l'innovation sur les données peut apporter à la statistique publique et sur le changement de paradigme que cela implique.

Considérations générales

Les considérations générales qui doivent guider l'OFS dans la mise en œuvre des étapes décrites plus haut (et dans tout ce qui touche à l'innovation sur les données) doivent être axées sur les valeurs centrales que sont la confidentialité des données, la protection des données, la sécurité des données, la gouvernance des données et la confiance du public.

Réalisation des objectifs stratégiques

Les actions suivantes seront nécessaires pour attendre les objectifs stratégiques précités et pour mettre en œuvre l'innovation sur les données:

- Évaluer la nécessité d'adapter et de développer le système qualité actuel de l'OFS pour y intégrer l'innovation sur les données (par exemple les méthodes d'analyse complémentaires telles que l'analyse prédictive);
- Évaluer les compétences nouvelles à faire acquérir au personnel de l'OFS dans le domaine des méthodes d'analyse complémentaires et des technologies (et outils) informatiques. La 1^{re} étape mentionnée plus haut (et les projets pilotes qui s'y rapportent) jouent un rôle clé pour déterminer les besoins en termes de compétences dans le domaine de l'innovation sur les données. Une fois cette évaluation réalisée, l'OFS pourra (si nécessaire):
 - développer avec des entreprises privées et/ou des universités de nouveaux cours de formation internes sur les méthodes d'analyse complémentaires («formation continue du personnel»), et développer d'autres compétences générales nécessaires: créativité, collaboration, curiosité, résolution de problèmes, communication et *storytelling* («compétences humaines»);
 - définir et mettre en place un «laboratoire de données» interne et/ou externe («*sandbox*») pour développer les aptitudes nouvelles nécessaires à la mise en œuvre des méthodes d'analyse complémentaires;
 - déterminer dans quelle mesure il est nécessaire d'adapter les technologies (informatiques) et les infrastructures et outils internes et/ou externes de l'OFS;
- Étudier les questions légales liées à l'innovation sur les données, p. ex. les dispositions et incitations légales nécessaires pour obtenir l'accord des entités privées susceptibles de communiquer des données secondaires («propriété de données»), sans compromettre les intérêts économiques des propriétaires des données (pour la 3^e étape ci-dessus);
- Discuter et échanger des informations avec toutes les parties concernées au sein de l'OFS, au sein du système statistique et dans la communauté des chercheurs.

L'OFS envisage de nouer des partenariats multidisciplinaires durables et efficaces entre la Confédération, l'industrie, les universités et la communauté statistique pour rassembler toutes les compétences et les expériences nécessaires à la réussite de sa stratégie d'innovation sur les données.

Commandes d'imprimés

Tél. 058 463 60 60

Fax 058 463 60 61

order@bfs.admin.ch

Prix

gratuit

Téléchargement

www.statistique.ch (gratuit)

Numéro OFS

1788-1700

La statistique www.la-statistique-compte.ch
compte pour vous.