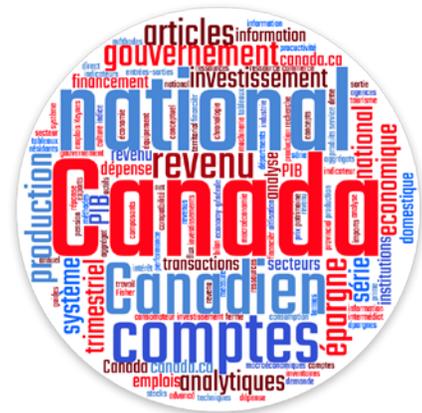


## Les nouveautés en matière de comptes économiques canadiens

### La valeur des données au Canada : estimations expérimentales



Date de diffusion : le 10 juillet 2019



Statistique  
Canada

Statistics  
Canada

Canada

## Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à [www.statcan.gc.ca](http://www.statcan.gc.ca).

Vous pouvez également communiquer avec nous par :

Courriel à [STATCAN.infostats-infostats.STATCAN@canada.ca](mailto:STATCAN.infostats-infostats.STATCAN@canada.ca)

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

- |   |                |
|---|----------------|
| • Service de renseignements statistiques                                    | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur   | 1-514-283-9350 |

### Programme des services de dépôt

- |                             |                |
|-----------------------------|----------------|
| • Service de renseignements | 1-800-635-7943 |
| • Télécopieur               | 1-800-565-7757 |

## Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site [www.statcan.gc.ca](http://www.statcan.gc.ca) sous « Contactez-nous » > « Normes de service à la clientèle ».

## Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Publication autorisée par le ministre responsable de Statistique Canada

© Sa Majesté la Reine du chef du Canada, représentée par le ministre de l'Industrie 2019

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'[entente de licence ouverte](#) de Statistique Canada.

Une [version HTML](#) est aussi disponible.

*This publication is also available in English.*

# La valeur des données au Canada : estimations expérimentales

## Introduction

Dans un document précédent<sup>1</sup>, on a examiné les récentes avancées dans le domaine de la collecte, de la numérisation, du stockage et de l'exploitation de l'information dans le monde entier. L'utilisation de l'information est généralisée et on la retrouve dans les lieux de travail, à la maison, dans les administrations publiques, dans les systèmes de communication et de transport et ailleurs. Pourtant, le phénomène n'apparaît pas encore de manière évidente dans les données économiques. Un élargissement des concepts et des méthodes de la comptabilité nationale a été proposé sous la forme d'une hiérarchie de l'information dans laquelle certaines « observations » de la vie quotidienne sont numérisées et deviennent des « données ». Ces données sont alors structurées et organisées dans des « bases de données » à des fins d'utilisation pratique. Puis, les chercheurs et les entreprises accèdent à ces bases de données et utilisent la « science des données » pour élaborer et tester des hypothèses et déboucher sur de nouvelles découvertes d'intérêt concernant le monde réel.

Le présent article élargit et, dans une certaine mesure, teste le cadre statistique exposé dans le document précédent et présente un ensemble provisoire d'estimations statistiques sur les montants investis dans les données, les bases de données et la science des données canadiens au cours des dernières années. Les estimations sont calculées à partir des renseignements sur l'emploi et les salaires recueillis lors du recensement quinquennal de la population et de l'enquête mensuelle sur la population active, lesquels sont combinés à un certain nombre d'hypothèses importantes mais qui n'ont pas encore été, dans une large mesure, testées. Les résultats indiquent une croissance rapide des investissements en données, en bases de données et en science des données et une accumulation importante de ces types de capital au fil du temps.

## Évaluer les données, les bases de données et la science des données

Comme il a été mentionné dans le précédent document, les données, les bases de données et la science des données peuvent être produites et utilisées par les entreprises pour leur compte propre ou vendues sur le marché. Les données, les bases de données et la science des données vendues sur le marché sont théoriquement évaluées au prix du marché (la valeur de l'opération). Idéalement, Statistique Canada mènerait une enquête auprès des entreprises canadiennes et obtiendrait d'elles des renseignements sur leurs ventes de données, de bases de données et de science des données sur le marché. À l'heure actuelle, Statistique Canada dispose de très peu d'information sur les ventes de données, de bases de données et de science des données sur le marché. Les données, les bases de données et la science des données utilisées pour compte propre sont évaluées au coût de production du produit, y compris un rendement du capital estimé. Puisque Statistique Canada ne possède pas d'information sur les ventes de données, de bases de données et de la science des données sur le marché, la production de toutes les données, bases de données et science des données, que ce soit pour la vente sur le marché ou pour utilisation pour compte propre, a été évaluée au coût de production du produit.

## Données

Les données, telles que définies dans le document cité précédemment, sont produites et donc incluses et évaluées dans les limites de la production du système de comptabilité nationale (SCN). Dans certains cas, les données sont achetées et vendues lors d'opérations sur le marché. Dans ces situations, la valeur est tout simplement le prix du marché. Dans d'autres cas peut-être plus courants, les données sont produites et utilisées au sein d'une entreprise, d'une administration publique ou d'une institution sans but lucratif. Dans ces cas, étant donné qu'une valeur déterminée par le marché sans lien de dépendance est indisponible, la valeur associée doit être estimée.

Ainsi, si une entreprise achète des données d'une autre entreprise, la valeur correspond au prix de l'opération. Par exemple, si Statistique Canada achète des renseignements financiers à Bloomberg Canada, les données seront évaluées au prix négocié entre les deux parties.

1. Statistique Canada, « [Mesurer les investissements en données, en bases de données et en science des données : cadre conceptuel](#) ».

Traditionnellement, la méthode utilisée pour évaluer les produits pour compte propre (créés et utilisés à l'interne) consistait à additionner les coûts liés à leur production, « majorés » par un rendement normal du capital. Comme indiqué précédemment, le coût de la numérisation des observations pour compte propre (à la marge) est souvent proche de zéro puisqu'il peut ne pas nécessiter d'intrant travail. Par exemple, cela devient de plus en plus le cas avec l'« Internet des objets » dans lequel des capteurs numérisent automatiquement des observations et les stockent dans une base de données<sup>2</sup>.

Les exemples d'activités liées à la production de données vont des coûts de la main-d'œuvre associés à la saisie de renseignements sur papier sous une forme lisible par une machine aux coûts associés à l'utilisation d'un drone pour obtenir des images numériques d'un emplacement géographique. En outre, les progrès de l'intelligence artificielle et de l'apprentissage automatique permettent de construire des algorithmes complexes en langage naturel, qui prennent des informations numériques non structurées (telles qu'une photo) et les transforment en informations codées et hautement structurées à partir desquelles des bases de données peuvent être créées et des connaissances peuvent être acquises.

Par exemple, Statistique Canada est responsable de la diffusion de statistiques sur le camionnage destinées aux Canadiens. Ces statistiques comprennent des renseignements relatifs à l'origine et à la destination des marchandises transportées. Compte tenu des centaines de milliers de livraisons qui ont lieu chaque jour, demander aux entreprises de camionnage de résumer tous ces renseignements dans le cadre d'une enquête leur impose un lourd fardeau. Afin d'alléger ce fardeau, l'agence a négocié l'acquisition de connaissances électroniques auprès d'un certain nombre d'entreprises de camionnage. Ces connaissances comprennent une quantité considérable de détails relatifs au produit, à son origine et à sa destination. Ces connaissances sont tous saisis numériquement par les entreprises de camionnage et les données sont transférées à Statistique Canada. Pour Statistique Canada, le défi consiste à prendre les descriptions non structurées de chaque connaissance et à les classer dans un système de codage de la classification des produits standards utilisé par l'agence lors de la production de statistiques sur le camionnage. Pour ce faire, l'agence a employé un certain nombre de spécialistes des données qui créent des algorithmes complexes afin de garantir la classification correcte des données. Le travail de ces scientifiques des données ferait partie de la « valeur marchande » imputée des données acquises par Statistique Canada à partir de ces connaissances électroniques afin de contribuer à l'élaboration d'un ensemble amélioré de statistiques sur le camionnage.

Dans ce document, la valeur des données est estimée en fonction des coûts de la main-d'œuvre liés à leur production, auxquels s'ajoutent les coûts indirects de la main-d'œuvre et d'autres coûts, tels que les coûts de gestion des ressources humaines et de contrôle financier, l'électricité, l'entretien des bâtiments et les services de télécommunications qui y sont associés.

Les groupes professionnels sont choisis parmi ceux de la Classification nationale des professions (CNP), généralement associés à la conversion des observations en format numérique (processus de numérisation). Les groupes professionnels engagés dans cette activité sont présentés dans le tableau 1.

Les employés travaillant dans ces catégories de la CNP ont peu de chances de passer tout leur temps à produire des données. Ils peuvent également être engagés dans plusieurs autres types d'activités. Les renseignements sur la part de leur temps de travail consacrée à la production de données étant actuellement indisponibles, des hypothèses subjectives ont été formulées. Compte tenu de l'incertitude associée à ces hypothèses, deux solutions ont été envisagées. Elles sont étiquetées « fourchette inférieure » et « fourchette supérieure » et figurent dans le tableau 1. Des travaux supplémentaires devront être effectués dans l'avenir pour recueillir des renseignements factuels sur les groupes professionnels spécifiques qui exercent des activités de production de données et sur les parts de leurs intrants travail associées à cette activité.

Comme il a été mentionné, deux sources de données sont utilisées pour cette étude. La première, le recensement quinquennal de la population, fournit des statistiques de qualité sur l'emploi et les gains selon la profession. Ces informations clés sont utilisées pour chacune des années de recensement 2006, 2011 et 2016 et elles se rapportent aux années précédant immédiatement les recensements. L'enquête mensuelle sur la population active est l'autre source de données utilisée; elle est plus fréquente et plus à jour que le recensement, mais moins précise

---

2. Ceci fait référence aux dispositifs et objets dont l'état peut être modifié à l'aide d'Internet, avec ou sans la participation active d'individus. OCDE (2015), *OECD Digital Economy Outlook 2015*, Éditions OCDE, Paris.

en raison de la taille relativement petite de son échantillon. Les données de cette enquête ont été utilisées pour les années autres que 2005, 2010 et 2015.

On suppose que les salaires non directs et les autres coûts représentent 50 % des dépenses salariales<sup>3</sup>. Une marge supplémentaire de 3 % est ajoutée à cette marge pour les services du capital. Ce modèle est similaire au modèle utilisé ailleurs dans les comptes nationaux canadiens pour mesurer la valeur des logiciels pour compte propre et les coûts d'investissement en recherche et développement.

**Tableau 1**  
**Investissement en « données »**

	Activités de production liées aux « données »	Investissement en « données »			
		2005 à 2018	2005	2010	2015
	pourcentage	millions de dollars			
<b>Total des groupes professionnels</b>					
valeur de fourchette inférieure	...	6 777	7 559	8 916	9 418
valeur de fourchette supérieure	...	9 742	10 840	13 448	14 216
<b>Analystes financiers et analystes en placements</b>					
valeur de fourchette inférieure	10	475	456	1 124	992
valeur de fourchette supérieure	20	949	913	2 249	1 983
<b>Superviseurs des services d'information et des services à la clientèle</b>					
valeur de fourchette inférieure	30	578	342	307	307
valeur de fourchette supérieure	50	964	569	511	512
<b>Commis à l'entrée des données</b>					
valeur de fourchette inférieure	100	2 041	2 114	1 924	1 942
valeur de fourchette supérieure	100	2 041	2 114	1 924	1 942
<b>Autres préposés aux services d'information et aux services à la clientèle</b>					
valeur de fourchette inférieure	30	2 534	2 901	3 517	3 576
valeur de fourchette supérieure	50	4 223	4 835	5 862	5 959
<b>Intervieweurs pour enquêtes et commis à la statistique</b>					
valeur de fourchette inférieure	90	409	541	419	215
valeur de fourchette supérieure	100	454	602	466	239
<b>Mathématiciens, statisticiens et actuaires</b>					
valeur de fourchette inférieure	20	165	325	398	930
valeur de fourchette supérieure	30	248	488	597	1 395
<b>Économistes, chercheurs et analystes des politiques économiques</b>					
valeur de fourchette inférieure	20	238	374	555	790
valeur de fourchette supérieure	30	357	562	832	1 184
<b>Rechercheurs, experts-conseils et agents de programmes en politiques sociales</b>					
valeur de fourchette inférieure	20	338	505	672	667
valeur de fourchette supérieure	30	507	757	1 008	1 000
		pourcentage			
<b>Taux de croissance annuel</b>					
valeur de fourchette inférieure	...	...	2,2	3,4	1,8
valeur de fourchette supérieure	...	...	2,2	4,4	1,9

... n'ayant pas lieu de figurer

**Note :** Le masculin est utilisé de façon générique et désigne aussi bien les hommes que les femmes.

**Source :** Statistique Canada, totalisations spéciales.

Le tableau 1 indique qu'en 2018, entre 9 milliards de dollars et 14 milliards de dollars ont été dépensés en formation brute de capital fixe pour les données. Il s'agit d'un montant relativement peu élevé lorsqu'on le compare au montant consacré à la formation brute de capital fixe totale au Canada, qui s'élevait à 498 milliards de dollars cette même année. Les taux de croissance annuels dans cette catégorie d'investissement sont également modestes (voir le tableau 1).

Comme indiqué précédemment, compte tenu des hypothèses retenues et du fait que la plupart des observations sont numérisées au moyen de processus automatisés, il n'est pas surprenant que même si des zettaoctets d'informations sont générés chaque année, cette valeur est relativement petite.

3. Statistique Canada détient quelque expérience dans le calcul des coûts indirects tels que ceux qui sont liés à ses propres programmes à coûts recouvrables. Statistique Canada facture ses clients pour les enquêtes, les tableaux et d'autres travaux à but spécifique personnalisés et les prix sont fondés sur les coûts directs majorés pour tenir compte des coûts indirects.

Bien que la valeur des investissements puisse être faible, l'importance économique et sociale des données augmente rapidement. L'ajout de ce type d'investissement aux états financiers des propriétaires économiques des données donne une plus grande importance à la catégorie de produit. Le fait d'inscrire les immobilisations correspondantes au bilan constitue un début de reconnaissance de ce que beaucoup considèrent comme une composante manquante importante du patrimoine.

En outre, la valeur des données reflète les coûts des intrants et non le flux de revenus potentiel pouvant être saisi à partir des données. Cela signifie que les 9 à 14 milliards de dollars susmentionnés constituent en soi une estimation à la limite inférieure, car ils ne tiennent pas compte de toutes les utilisations potentielles futures des données. Des recherches supplémentaires et la mise au point de méthodes appropriées sont nécessaires avant de pouvoir établir des estimations crédibles du futur flux de revenus tiré des données.

## Base de données

Les méthodes recommandées pour évaluer les bases de données sont décrites dans le manuel du Système de comptabilité nationale de 2008 (SCN 2008). Il y est noté que la valeur d'une base de données devra généralement être estimée selon la méthode de la somme des coûts (paragraphe 10.113). Les coûts comprennent :

- coût de préparation des données dans le format approprié pour la base de données;
- temps passé par le personnel à élaborer la base de données;
- les services du capital des actifs utilisés dans l'élaboration de la base de données;
- les coûts des articles utilisés en consommation intermédiaire.

Les bases de données achetées sur le marché devraient être évaluées aux prix d'acquisition, alors que celles élaborées à l'interne devraient l'être à leur prix de base estimé ou à leurs coûts de production (y compris le rendement du capital des producteurs marchands) s'il n'est pas possible d'estimer le prix de base (paragraphe A3.60).

Dans la plupart des cas, le défi des responsables de la comptabilité nationale n'est pas un problème conceptuel mais plutôt un problème de « manque d'information ». Les frontières entre les logiciels, les bases de données et les services (tels que les services d'assistance aux clients) sont souvent floues. Par conséquent, dans de nombreux cas, les organismes statistiques supposent que les investissements en bases de données sont pris en compte dans les estimations de la formation brute de capital fixe en logiciels. Statistique Canada fait la même chose, même s'il est prouvé que les bases de données ne sont pas entièrement saisies avec la méthodologie actuelle.

Pour les besoins de cet article, Statistique Canada a élaboré une méthodologie pour estimer la valeur des bases de données pour compte propre séparément des logiciels.

La méthodologie existante de Statistique Canada pour estimer les investissements en logiciels pour compte propre (y compris les bases de données) consiste à identifier un certain nombre de groupes professionnels liés à l'élaboration de logiciels et de bases de données et à formuler des hypothèses sur le temps que ces groupes d'employés passent à élaborer des logiciels et des bases de données pour utilisation finale pour compte propre dans l'entreprise. Outre le coût de l'intrant travail, Statistique Canada inclut également les coûts non liés à la main-d'œuvre associés à l'élaboration de logiciels tels que l'électricité, la location d'immeubles et d'autres types de frais généraux.

Actuellement, Statistique Canada combine les activités des groupes professionnels<sup>4</sup> suivants dans son calcul de l'investissement en logiciels pour compte propre :

- 2171 Analystes et consultants en informatique
- 2172 Analystes de bases de données et administrateurs de données
- 2173 Ingénieurs et concepteurs de logiciel
- 2174 Programmeurs et développeurs en médias interactifs
- 2175 Concepteurs et développeurs Web

---

4. Le masculin est utilisé de façon générique et désigne aussi bien les hommes que les femmes.

La première étape pour séparer les investissements en logiciels pour compte propre (y compris les bases de données) des investissements en bases de données pour compte propre consiste à distinguer les groupes professionnels mentionnés ci-dessus qui sont liés aux logiciels pour compte propre et ceux qui sont liés aux bases de données pour compte propre. Pour les besoins de cette étude, on suppose que le groupe professionnel 2172 de la CNP concerne les bases de données et que les autres concernent l'élaboration de logiciels.

En plus d'inclure le groupe professionnel 2172 de la CNP dans la production de bases de données, d'autres groupes professionnels ont été examinés afin de déterminer si certaines de leurs activités sont liées à l'élaboration de bases de données. Il a été décidé que, aux fins d'estimation des investissements en bases de données, une partie de l'intrant travail des groupes professionnels suivants serait également incluse :

- 0213 Gestionnaires des systèmes informatiques
- 2283 Évaluateurs de systèmes informatiques

Lors du calcul de l'évolution des investissements en logiciels et en bases de données pour compte propre, des hypothèses doivent encore être faites concernant la proportion de l'activité d'un employé devant être capitalisée. En moyenne, les personnes élaborant des logiciels ne consacrent pas 100 % de leur temps à l'élaboration de logiciels internes. Une partie de leur temps peut être consacrée à l'élaboration de logiciels standards vendus à d'autres entreprises ou à des consommateurs, par exemple.

Comme c'est le cas pour les estimations des investissements en données, il est nécessaire de faire des hypothèses pour les estimations des investissements en bases de données concernant la part de leur temps que les employés consacrent à cette activité, selon le groupe professionnel. Les valeurs inférieures et supérieures pour ces hypothèses sont présentées dans le tableau 2 de même que les estimations de la valeur de la production de bases de données pour compte propre.

**Tableau 2**  
**Investissement en « bases de données »**

	Activités de production liées aux « bases de données »		Investissement en « bases de données »			
	2005 à 2018		2005	2010	2015	2018
	pourcentage		millions de dollars			
<b>Total des groupes professionnels</b>						
valeur de fourchette inférieure	...		3 087	4 143	5 945	8 046
valeur de fourchette supérieure	...		4 564	6 104	8 599	11 625
<b>Gestionnaires de systèmes informatiques et d'information</b>						
valeur de fourchette inférieure	30	1 880	2 527	3 345	4 555	
valeur de fourchette supérieure	50	3 133	4 211	5 574	7 591	
<b>Analystes de bases de données et administrateurs de données</b>						
valeur de fourchette inférieure	90	1 045	1 444	2 357	3 212	
valeur de fourchette supérieure	100	1 162	1 604	2 619	3 569	
<b>Évaluateurs de systèmes informatiques</b>						
valeur de fourchette inférieure	30	161	173	244	279	
valeur de fourchette supérieure	50	269	289	406	466	
			pourcentage			
<b>Taux de croissance annuel</b>						
valeur de fourchette inférieure	...	...	6,1	7,5	10,6	
valeur de fourchette supérieure	...	...	6,0	7,1	10,6	

... n'ayant pas lieu de figurer

**Note :** Le masculin est utilisé de façon générique et désigne aussi bien les hommes que les femmes.

**Source :** Statistique Canada, totalisations spéciales.

La valeur totale des investissements en bases de données pour compte propre en 2018 est estimée entre 8 milliards de dollars et 12 milliards de dollars. De cette valeur, environ un milliard de dollars représente une réaffectation de logiciels à des bases de données.

Les taux estimés de la croissance des investissements en bases de données (voir le tableau 2) ont été très élevés et à la hausse au cours des dernières années. Ils étaient de 6 % par année de 2005 à 2010, de 7 % par année de 2010 à 2015 et de 10,5 % par année de 2015 à 2018.

À l'heure actuelle, Statistique Canada ne recueille pas d'information sur les achats de bases de données sur le marché. En principe, une base de données peut être vendue plusieurs fois, mais les renseignements sur les ventes faites par cette industrie ne sont pas disponibles. En conséquence, aux fins de la présente étude, seule la production de bases de données pour compte propre est incluse dans les calculs des actifs liés aux bases de données.

## Science des données

Tout comme les bases de données, le manuel du SCN 2008 fournit aux responsables de la comptabilité nationale une méthode standard pour estimer la valeur des investissements en recherche et développement. Lorsque les résultats de la recherche et développement sont vendus sur le marché, le prix du marché est utilisé pour l'évaluation. Lorsque la recherche et développement est entreprise pour une utilisation finale pour compte propre, une approche de la somme des coûts est utilisée. Dans le cas du Canada, alors que les analyses de données sont incluses dans la recherche et développement en principe et que le cadre conceptuel et les méthodes pour les mesurer existent, le nombre croissant d'entreprises engagées dans l'analyse de données signifie qu'il existe une sous-estimation statistique potentielle. Le problème est que les instruments de collecte actuels sont conçus pour recueillir des renseignements auprès d'un ensemble relativement restreint d'entreprises qui sont connues pour être axées sur la recherche.

Les estimations des activités de recherche et développement au Canada reposent sur deux sources principales d'information. L'Enquête annuelle sur la recherche et le développement dans l'industrie canadienne (RDIC) sert à mesurer les activités de recherche et développement des entreprises des secteurs non financier et financier. Il s'agit d'une enquête transversale portant sur environ 8 000 entreprises. Un certain nombre d'enquêtes menées par les administrations publiques fédérale et provinciales sont utilisées pour mesurer les activités de recherche et développement dans le secteur des administrations publiques.

Bien que les stratégies d'échantillonnage et d'enquête de l'enquête sur la RDIC conviennent aux formes traditionnelles de recherche et développement, telles que la recherche et développement pharmaceutique ou le génie logiciel, elles ne sont pas aussi bien conçues, que ce soit en termes d'instrument ou d'échantillonnage, pour rendre compte de la recherche croissante utilisant les mégadonnées, qui est appelée la science des données. Par exemple, les détaillants et les banques utilisent les informations tirées de leurs immenses réserves de données personnelles pour accroître leurs ventes. Ces informations correspondent à la définition de la recherche et développement du manuel du SCN 2008. Le problème, du moins dans le cas du Canada, est que les méthodes et les outils statistiques actuels ne rendent pas pleinement compte de cette activité d'investissement dans la recherche et développement.

Afin d'élaborer une estimation par ordre de grandeur de la valeur des investissements en science des données au Canada, la même approche que celle décrite ci-dessus pour les données et les bases de données est adoptée. Il est supposé que les activités de la science des données se produisent dans les groupes professionnels indiqués dans le tableau 3. Une part des activités de production est également supposée pour chacun de ces groupes et, comme pour les données et les bases de données, une majoration hypothétique pour les coûts non directs de la main-d'œuvre et les autres coûts est appliquée aux coûts directs de la main-d'œuvre.

**Tableau 3**  
**Investissement en « science des données »**

	Activités de production liées à la « science des données »	Investissement en « science des données »			
		2005 à 2018	2005	2010	2015
	pourcentage	millions de dollars			
<b>Total des groupes professionnels</b>					
valeur de fourchette inférieure	...	4 829	6 085	11 168	11 991
valeur de fourchette supérieure	...	5 689	7 181	13 145	14 184
<b>Analystes financiers et analystes en placements</b>					
valeur de fourchette inférieure	60	2 848	2 738	6 746	5 950
valeur de fourchette supérieure	70	3 323	3 194	7 870	6 942
<b>Agents de statistiques et professions connexes du soutien de la recherche</b>					
valeur de fourchette inférieure	90	129	336	360	76
valeur de fourchette supérieure	100	144	373	400	84
<b>Mathématiciens, statisticiens et actuaires</b>					
valeur de fourchette inférieure	50	413	813	995	2 324
valeur de fourchette supérieure	60	495	976	1 194	2 789
<b>Économistes, recherchistes et analystes des politiques économiques</b>					
valeur de fourchette inférieure	50	595	936	1 387	1 974
valeur de fourchette supérieure	60	714	1 123	1 664	2 369
<b>Recherchistes, experts-conseils et agents de programmes en politiques sociales</b>					
valeur de fourchette inférieure	50	844	1 262	1 681	1 667
valeur de fourchette supérieure	60	1 013	1 514	2 017	2 000
		pourcentage			
<b>Taux de croissance annuel</b>					
valeur de fourchette inférieure	...	...	4,7	12,9	2,4
valeur de fourchette supérieure	...	...	4,8	12,9	2,6

... n'ayant pas lieu de figurer

**Note :** Le masculin est utilisé de façon générique et désigne aussi bien les hommes que les femmes.

**Source :** Statistique Canada, totalisations spéciales.

Les activités liées à la « science des données » sont estimées pour rendre compte de la formation brute de capital fixe d'entre 12 milliards de dollars et 14 milliards de dollars en 2018. Les taux de croissance annuels dans cette catégorie d'investissement ont également été considérables : près de 5 % de 2005 à 2010, environ 13 % de 2010 à 2015 et près de 2,5 % de 2015 à 2018.

## Investissements totaux dans les actifs liés aux données

Le tableau 4 présente la formation brute totale de capital fixe dans chaque composante de la chaîne d'information pour 2018. L'investissement total s'élève entre 29 milliards de dollars et 40 milliards de dollars aux prix courants. La croissance de 2005 à 2018 est de 100 %, soit 5,5 % sur une base annuelle moyenne. Ces montants ne peuvent toutefois pas être ajoutés aux estimations existantes du produit intérieur brut, car ils chevauchent dans une certaine mesure les estimations publiées de la formation brute totale de capital fixe. Des travaux supplémentaires sont nécessaires pour calculer le chevauchement et affiner les estimations.

**Tableau 4**  
**Investissement en « données », « bases de données » et « science des données »**

	2005	2010	2015	2018
	millions de dollars			
<b>Total des catégories liées aux données</b>				
valeur de fourchette inférieure	14 693	17 788	26 029	29 455
valeur de fourchette supérieure	19 995	24 125	35 192	40 025
<b>« Données »</b>				
valeur de fourchette inférieure	6 777	7 559	8 916	9 418
valeur de fourchette supérieure	9 742	10 840	13 448	14 216
<b>« Bases de données »</b>				
valeur de fourchette inférieure	3 087	4 143	5 945	8 046
valeur de fourchette supérieure	4 564	6 104	8 599	11 625
<b>« Science des données »</b>				
valeur de fourchette inférieure	4 829	6 085	11 168	11 991
valeur de fourchette supérieure	5 689	7 181	13 145	14 184
	pourcentage			
<b>Taux de croissance annuel</b>				
valeur de fourchette inférieure	...	3,9	7,9	4,2
valeur de fourchette supérieure	...	3,8	7,8	4,4

... n'ayant pas lieu de figurer

Source : Statistique Canada, totalisations spéciales.

## Le stock d'actifs liés aux données

Ce document suggère qu'une quantité importante d'activités « liées à l'information » crée des réserves de valeur, auxquelles les entreprises puisent au cours de périodes ultérieures pour produire des biens et des services. Étant donné que les entreprises, les administrations publiques et les institutions sans but lucratif investissent chaque jour dans les données, les bases de données et la science des données pour compte propre, un stock de ces actifs est également constitué. Ce stock doit être inclus dans le bilan du secteur qui en est propriétaire, à sa valeur marchande.

Trois méthodes peuvent être utilisées pour mesurer le stock de données, les bases de données et la science des données. Une méthode consisterait à traiter l'actif de la même manière qu'une ressource naturelle et à actualiser le futur flux de revenus pouvant être généré à partir de la chaîne d'information. Le problème avec cette approche est que les avantages potentiels (et donc le flux de revenus) ne sont jamais connus avec certitude. Étant donné que les données peuvent avoir de nombreuses utilisations et que les mêmes données peuvent avoir de multiples utilisations, les revenus potentiels sont illimités. Dans le cas d'une ressource naturelle, le stock de la ressource, les utilisations, le modèle d'utilisation, le prix et la durée jusqu'à l'épuisement du stock connu sont largement compris. Dans le cas des données, le temps qu'elles seront utilisées est inconnu, tout comme le sont le prix (car cela dépend de l'utilisation) et les utilisations potentielles. Bien qu'il soit vrai que certaines de ces informations sont prises en compte dans la capitalisation boursière de grandes entreprises de traitement de données ou de grandes banques de données, elles sont sujettes à d'importantes fluctuations et elles seraient difficiles à utiliser pour élaborer une estimation du volume de stock de données, de bases de données et de science des données. Par conséquent, aux fins du présent document, cette approche n'est pas adoptée puisque davantage de recherche est nécessaire pour élaborer des estimations crédibles.

Une seconde approche consiste à indiquer la valeur des données, des bases de données et de la science des données comme elle est enregistrée dans les états financiers des entreprises. En théorie, cette valeur devrait refléter la valeur marchande de l'actif et incorporer, autant que faire se peut, le volume des données reflété au prix de vente éventuel courant. Malheureusement, très peu d'entreprises enregistrent directement les données, les bases de données et la science des données dans leurs bilans. Étant donné la nature intangible de l'actif, il est soit non enregistré, soit combiné avec d'autres types d'actifs intangibles comme la bonne volonté.

Comme troisième approche, le manuel du SCN 2008 recommande l'utilisation de la méthode de l'inventaire permanent (MIP) pour l'établissement d'estimations d'un stock d'actifs. La MIP accumule les flux d'investissement dans le temps, aux prix constants, assume une fonction d'élimination et un profil d'amortissement et intègre les niveaux de prix actuels pour calculer une estimation de la valeur marchande de la valeur du stock d'un actif.

Évaluer le stock de données, les bases de données et les actifs liés à la science des données soulève un certain nombre de défis intéressants. Le premier concerne le profil d'amortissement. Ces actifs ne se déprécient pas physiquement, un profil naturellement observable ne peut donc pas être utilisé. Dans certains cas, les entreprises stockent des quantités massives d'informations indéfiniment, bien que l'utilité perçue de ces actifs diminue. Parallèlement, dans d'autres cas, la valeur de l'information est éphémère et ne devrait pas être capitalisée si elle n'est pas utilisée pendant une période supérieure à un an (savoir qu'il pleuvrait hier était plus utile avant-hier et hier).

Pour des raisons pragmatiques, il est supposé que les données ont une durée de vie utile de 25 ans, les bases de données ont la même durée de vie que les logiciels, soit 5 ans, et les résultats de la recherche et développement axée sur les données ont une durée de vie de 6 ans, la même hypothèse étant utilisée pour d'autres formes de recherche et développement.

La raison de l'hypothèse d'une durée de vie utile de 25 ans pour les données est basée sur la durée pendant laquelle on s'attend à ce qu'une entreprise stocke des données ou utilise au moins des données stockées pour mieux comprendre certaines choses. Étant donné que beaucoup de données actuellement utilisées sont comportementales, on peut supposer que ces données ne conserveront leur valeur que pour une « génération ». Une génération est souvent définie comme la période nécessaire pour que les enfants naissent, grandissent, deviennent adultes et commencent à avoir leurs propres enfants. Bien sûr, il existe de nombreux autres types de données. L'hypothèse de 25 ans doit être considérée comme assez provisoire et des recherches supplémentaires sont nécessaires concernant ce sujet.

Un stock de capital net avec un profil de dépréciation géométrique a été estimé pour les trois types d'actifs liés aux données.

Le deuxième défi lié à la mesure du stock de données est l'établissement d'un « prix du marché actuel ». Dans les sections précédentes, une approche de l'évaluation des données est décrite, mais cette approche ne s'applique qu'à la valeur initiale des données. La valeur marchande des données peut varier considérablement d'une période à l'autre.

Supposons que, s'appuyant sur le quatrième exemple du premier document, la société SearchBook a accumulé des données pendant des années mais qu'elle a simplement stocké ces données sans entreprendre aucune activité de recherche et développement qui y serait liée. Supposons encore qu'une autre société a créé un logiciel qui générera automatiquement des listes de courses pour les consommateurs, mais qui nécessite des renseignements historiques sur les habitudes d'achat. Le développeur de l'application, qui souhaite vendre l'application moyennant un abonnement de 5 \$ par mois, accepte d'acheter les données auprès de SearchBook pour un montant de 200 millions de dollars, alors que celles-ci n'avaient été évaluées qu'à 5 millions de dollars au bilan. Depuis qu'une nouvelle utilisation des données a été découverte, leur valeur augmente considérablement, même si aucune nouvelle donnée n'a été produite. Cet effet de réévaluation doit être reflété dans la valeur marchande de l'actif au bilan et dans le compte des autres changements d'actifs.

En général, lorsque que l'on utilise la MIP pour évaluer les stocks, il est nécessaire d'avoir des renseignements sur le prix du produit (données, bases de données et résultats de la science des données dans ce cas). Ces prix peuvent être utiles quand il s'agit de déflater une valeur marchande pour les données dans les cas où de nouvelles utilisations des données ne sont pas encore découvertes. Le problème est que les « nouvelles utilisations » des produits liés aux données qui sont souvent découvertes sont spécifiques à chaque produit et peuvent fondamentalement altérer la valeur. En tant que tel, l'approche de la somme des coûts en matière d'évaluation doit être accompagnée d'ajustements pour tenir compte des situations dans lesquelles il existe d'importantes réévaluations liées aux données. Les statisticiens ont besoin de travailler davantage pour obtenir des estimations de la valeur marchande directement observables, bien qu'imparfaites, dans certains cas. Les réévaluations *ex poste* de ce type ne sont pas prises en compte dans les estimations présentées dans le présent document.

Comme il est supposé que la plupart des données, des bases de données et de la science des données décrites dans le présent document sont produites et utilisées à l'interne par les entreprises, les administrations publiques et les organisations sans but lucratif, le prix de ces actifs dépendra du coût des intrants lié à la gestion directe, à la rémunération du travail et à la rémunération du travail non directe et aux coûts autres que de main-d'œuvre

tels que les services publics, les services de soutien aux employés et les services du capital. Aux fins du présent document, seuls les coûts directs du facteur travail ont été pris en compte lors de l'estimation des prix des données, des bases de données et de la science des données.

**Tableau 5**  
**Indice des prix de capital et stock de capital net géométrique, « données », « bases de données » et « science des données »**

	2005	2010	2015	2018
	2005=100			
<b>Total des indices des prix pour les catégories liées aux données</b>	100,0	109,5	119,1	126,3
« Données »	100,0	112,6	122,0	130,7
« Bases de données »	100,0	103,6	113,3	121,7
« Science des données »	100,0	108,4	116,9	121,2
	millions de dollars			
<b>Total du stock de capital net pour les catégories liées aux données</b>				
valeur de fourchette inférieure	74 058	100 512	131 950	157 067
valeur de fourchette supérieure	97 855	136 055	181 098	217 659
<b>« Données »</b>				
valeur de fourchette inférieure	53 549	74 181	92 133	104 824
valeur de fourchette supérieure	71 571	102 231	130 569	150 993
<b>« Bases de données »</b>				
valeur de fourchette inférieure	6 926	9 302	13 015	18 692
valeur de fourchette supérieure	10 290	13 740	18 954	27 050
<b>« Science des données »</b>				
valeur de fourchette inférieure	13 582	17 029	26 801	33 551
valeur de fourchette supérieure	15 993	20 084	31 576	39 616
	pourcentage			
<b>Taux de croissance annuel du total du stock de capital net</b>				
valeur de fourchette inférieure	...	6,3	5,6	6,0
valeur de fourchette supérieure	...	6,8	5,9	6,3

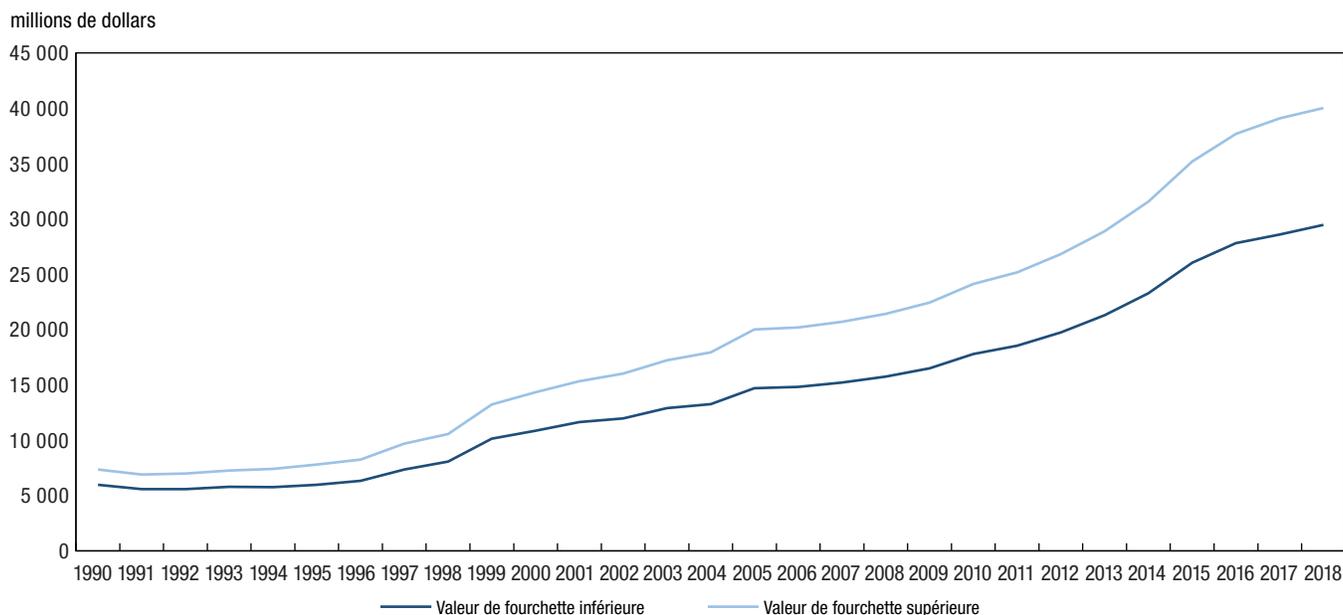
... n'ayant pas lieu de figurer

Source : Statistique Canada, totalisations spéciales.

Le tableau 5 présente les indices de prix des données, des bases de données et de la science des données en fonction du coût des intrants de chacune de ces composantes et du prix du service du capital supposé de 3 %. Il présente également les stocks nets de capital estimés à la fin des quatre années 2005, 2010, 2015 et 2018, aux prix courants, et les taux de croissance annuels moyens de ces estimations du stock de capital.

Bien que les résultats ne soient rapportés que pour quatre années récentes dans le tableau 5, la méthode a été appliquée chaque année à partir de 1990, en utilisant les statistiques sur l'emploi et la rémunération établies par la catégorie de la CNP de l'Enquête sur la population active. Les estimations des investissements pour cette plus longue période sont présentées dans le graphique 1. Les indices de prix sont calculés en supposant que les prix des investissements varient proportionnellement aux taux de rémunération du travail pour chaque CNP, ajustés à la baisse de 1 % par an pour tenir compte de la croissance supposée de la productivité. Comme indiqué précédemment, un profil de dépréciation géométrique est estimé avec une durée de vie utile de 25 ans pour les « données », de cinq ans pour les « bases de données » et de six ans pour la « science des données ». Les rejets sont supposés être nuls.

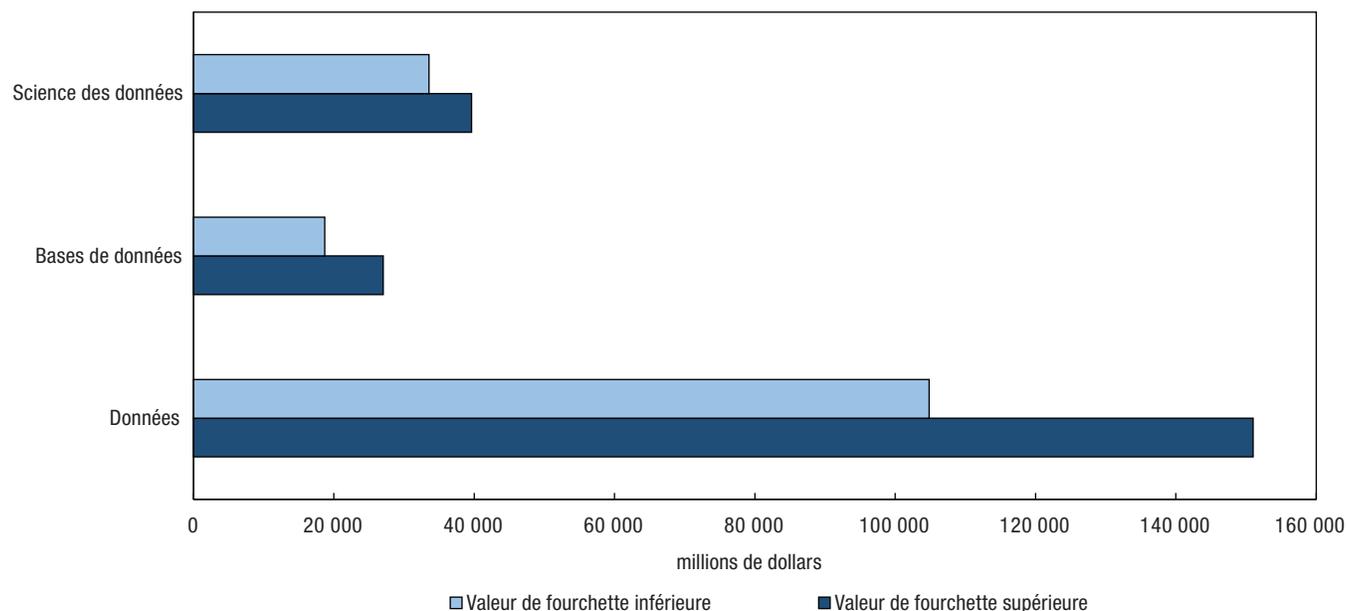
**Graphique 1**  
**Fourchette des investissements en les produits liés aux données, 1990 à 2018**



Source : Statistique Canada.

Les résultats indiquent un stock de capital net entre 157 milliards de dollars et 217 milliards de dollars à la fin de 2018. Les « données » représentent entre 105 milliards de dollars et 151 milliards de dollars de ce montant, les « bases de données », entre 19 milliards de dollars et 27 milliards de dollars et la recherche en « science des données », entre 34 milliards de dollars et 40 milliards (voir le graphique 2). Ces chiffres sont considérables, même s'ils sont faibles par rapport au total des capitaux de la construction non résidentielle, des machines et du matériel et des produits de la propriété intellectuelle, qui s'élèvent ensemble à 2 589 milliards de dollars. Le stock net pour les seuls produits de la propriété intellectuelle s'élève à 228 milliards de dollars.

**Graphique 2**  
**Fourchette du stock net de capital en les produits liés aux données, 2018**



Source : Statistique Canada.

Enfin, le tableau 6 présente les statistiques relatives aux investissements et au stock de capital net pour 2018, ventilées par secteur institutionnel. Le secteur des sociétés non financières représente environ la moitié des investissements et des stocks, tandis que les sociétés financières représentent un peu moins du tiers des investissements et le quart des stocks. Le secteur des administrations publiques représente environ un cinquième des investissements et des stocks, tandis que le secteur des institutions sans but lucratif au service des ménages représente environ 1,5 %.

**Tableau 6**  
**Investissement et stock de capital, « données », « bases de données », « sciences des données », par secteur, 2018**

	Total	Sociétés non financières		Sociétés financières		Secteur des administrations publiques		Institutions sans but lucratif au service des ménages		
		millions de dollars	millions de dollars	pourcentage du total	millions de dollars	pourcentage du total	millions de dollars	pourcentage du total	millions de dollars	pourcentage du total
<b>Investissement</b>										
valeur de fourchette inférieure	29 455	13 676	46,4	9 327	31,7	6 027	20,5	425	1,4	
valeur de fourchette supérieure	40 025	19 403	48,5	12 224	30,5	7 842	19,6	556	1,4	
<b>Stock de capital</b>										
valeur de fourchette inférieure	157 067	80 875	51,5	38 835	24,7	34 834	22,2	2 524	1,6	
valeur de fourchette supérieure	217 659	114 562	52,6	54 097	24,9	45 646	21,0	3 354	1,5	

Source : Statistique Canada, totalisations spéciales.

## Conclusions

La science des données et ses antécédents, les données et les bases de données, deviennent de plus en plus proéminentes dans le monde moderne. Une grande partie de ce que nous faisons de nos jours est numérisée sous forme de données, chargée dans des bases de données et exploitée de manière analytique à des fins très diverses. Pendant la journée, nos achats, nos voyages, nos lectures, notre écoute de musique, nos visionnements de contenu médiatique, nos activités physiques, les choses que l'on aime et celles que l'on n'aime pas et bien plus encore, sont stockés pour être utilisés à diverses fins. Même nos états physiques pendant le sommeil sont de plus en plus enregistrés<sup>5</sup>.

Cet état de fait nous est apparu assez rapidement dans le cours de l'histoire récente. Alors que les ordinateurs de bureau sont apparus dans les magasins dans les années 1970, Internet n'a commencé à être largement accessible qu'au milieu des années 1990. En 2002, le téléphone intelligent BlackBerry a été commercialisé pour la première fois par Research in Motion, une société fondée à Waterloo, en Ontario. Il offrait des services de téléphonie mobile, de courrier électronique, de navigation Web et d'autres services facilement accessibles sur un appareil sans fil de poche. Cinq ans plus tard, le premier iPhone a été annoncé par Steve Jobs, président-directeur général d'Apple Inc. Il offrait des fonctionnalités similaires à celles du BlackBerry, mais avec une interface utilisateur à écran tactile qui est rapidement devenue très populaire. Les années qui ont suivi ont été marquées par des améliorations considérables de la technologie iPhone, ainsi que par l'apparition de nombreux appareils Android concurrents et par d'importantes mises à niveau des réseaux de télécommunications. Ces développements technologiques et d'autres également, et leur adoption sans réserve par les Canadiens, ont facilité la collecte de vastes quantités de données.

Ces changements ont été rapides et le système statistique a du retard à rattraper. En effet, tous les organismes statistiques nationaux du monde entier sont actuellement confrontés à ce défi. Le présent document et celui qui l'a précédé, publié dans [Le Quotidien du 24 juin 2019](#) ont pour objet de faire un premier pas dans cette direction. Ils comportent un certain nombre d'hypothèses qui doivent être vérifiées. Les estimations numériques sont donc provisoires et présentées sous forme de fourchettes plutôt que d'évaluations ponctuelles. Néanmoins, les estimations indiquent des dépenses d'investissement et des stocks de capital significatifs et en hausse en ce qui a trait aux données, aux bases de données et à la science des données. Davantage de travail est à la fois justifié et nécessaire.

5. Voir, par exemple, « The 7 best sleep trackers of 2019 », [verywell fit](#).