

## ENQUÊTE SUR L'ACCÈS ET L'UTILISATION D'INTERNET AU QUÉBEC - UNE COLLECTE DE DONNÉES HYBRIDE -

Éric Lacroix<sup>1</sup>

### RÉSUMÉ

Au printemps 1998, le Bureau de la statistique du Québec (BSQ), en collaboration avec le Centre francophone de recherche en informatisation des organisations (CEFRIO) et le Réseau Interordinateurs scientifique québécois (RISQ), a mené une enquête qui a pour objet d'étudier les habitudes et les caractéristiques des internautes québécois. On peut concevoir la difficulté d'établir une base de sondage qui couvrirait de façon exhaustive la population visée : plusieurs internautes ont de multiples adresses de courrier électronique ou n'en ont aucune, les fournisseurs de connectivité sont réticents à donner les coordonnées de leur abonnés, certains québécois font affaire avec des fournisseurs situés à l'extérieur du Québec, etc.

Cet article résume les principales caractéristiques de la méthodologie de l'enquête qui a été retenue pour contourner ce problème, soit une combinaison du sondage téléphonique probabiliste traditionnel et de collecte des données sur le World Wide Web (WWW). On y discute, entre autres, des efforts requis pour assurer le suivi des répondants – que tous les internautes sélectionnés répondent ... une seule fois –, des aspects de la confidentialité, etc.

MOTS-CLÉS : Internet; collecte de données; WWW; questionnaire autoadministré.

### ABSTRACT

In the spring of 1998, the Bureau de la statistique du Québec (BSQ), in association with the Centre francophone de recherche en informatisation des organisations (CEFRIO) and the Réseau Interordinateurs scientifique québécois (RISQ) conducted a survey studying the practices and characteristics of Internet users in Quebec. One can easily understand the complexity involved in creating a survey frame that would present extensive coverage of this target population : many Internet users have multiple email addresses; some do not have email addresses; the ISPs (Internet Service Providers) are sometimes reluctant to give away the names and addresses of their customers; some Internet users are customers of ISFs who operate outside Quebec; etc.

This paper focuses on the characteristics of the survey methodology that was used to overcome this problem, a combination of traditional RDD telephone sampling and Web-based data collection. Among other things, the follow-up procedures are discussed –to get all the selected Internet users to complete the questionnaire...only once- as well as the confidentiality aspects surrounding the data collection.

KEY WORDS : Internet; Data Collection; WWW; Self-Administered Questionnaire.

## 1. INTRODUCTION

### 1.1 Contexte

Mis sur pied aux États-Unis à la fin des années 60 pour assurer la communication entre les différents centres de recherche de l'armée américaine, le réseau Internet est ensuite devenu le point de rencontre de la communauté universitaire internationale. Depuis quelques années, on a assisté à la démocratisation et à la commercialisation de ce nouveau média. Au

Québec, en particulier, le « réseau des réseaux » est accessible au grand public depuis 1989.

À l'ère de la société de l'information, l'attrait d'un tel outil est indéniable, mais on n'en saisit pas encore toutes les possibilités et il est difficile d'en percevoir les limites. D'un côté, on entend des discours euphoriques sur les possibilités de croissance exponentielle du commerce électronique, alors que d'autres constatent que la rentabilité n'est pas au rendez-vous. Dans toute cette effervescence, sans

<sup>1</sup> Eric Lacroix, Bureau de la statistique du Québec, 200 Chemin Sainte-Foy, Québec, Québec, Canada  
eric.lacroix@bsq.gouv.qc.ca

doute caractéristique de l'apparition d'un nouveau média de communication, les sondages et enquêtes sur le sujet se multiplient, tout le monde se vantant de connaître les internautes québécois, leurs habitudes, leurs besoins et leurs désirs. C'est dans ce contexte que s'inscrit l'enquête dont il est question dans le présent article.

Le Bureau de la statistique du Québec (BSQ) et le Centre francophone d'informatisation des organisations (CEFRIO) mènent conjointement depuis quelques années des enquêtes relatives à l'informatisation ou plus généralement aux technologies de l'information. Ainsi, en 1992, une première enquête auprès des ménages québécois avait permis de révéler que 11 p. cent possédaient un micro-ordinateur à la maison. En 1995, une autre enquête d'envergure s'intéressait aux entreprises et à leur niveau d'appropriation des technologies de l'information. Enfin, en novembre 1996, une nouvelle enquête auprès des ménages venait mettre à jour la mesure des taux d'informatisation domestique au Québec et s'intéressait, par ricochet, à l'utilisation du réseau Internet dans les foyers québécois.

De son côté, depuis mars 1996, le Réseau interordinateurs scientifique québécois (RISQ) menait des enquêtes à participation volontaire deux fois par année, dont l'objectif était de mesurer les habitudes des québécois sur Internet. La cueillette des données pour ces enquêtes, administrée à l'aide d'un questionnaire disponible sur le Web, durait un mois au cours duquel les internautes québécois étaient sollicités à partir de liens hypertextes présents dans les sites Web les plus fréquentés. La technologie de conception de questionnaires développée au RISQ permet l'utilisation de questions-filtres, et par conséquent l'administration de questionnaires complexes sur Internet.

Ainsi, ces trois organismes se sont concertés pour la réalisation d'une enquête probabiliste<sup>2</sup> auprès des internautes qui allierait la technologie de cueillette des données sur le Web développée par le RISQ, l'apport méthodologique du BSQ et l'expertise du CEFRIO dans le domaine des technologies de l'information.

---

<sup>2</sup> L'enquête a été rendue possible grâce à la contribution de partenaires publics : le ministère de la Culture et des Communications du Québec et le ministère des Relations avec les citoyens et de l'Immigration du Québec; et de partenaires privés : Bell Canada, Québec Téléphone et la Confédération des caisses populaires et d'économie Desjardins du Québec

## 1.2 Présentation de l'article

Le présent article met l'accent sur les aspects méthodologiques de l'enquête, particulièrement ceux qui concernent la collecte des données sur le réseau Internet. La description des objectifs de mesure sera suivie d'une brève présentation du contexte méthodologique de l'enquête. La section sur le questionnaire permettra d'introduire la collecte sur le Web, son fonctionnement, le suivi qui a été fait auprès des répondants et les résultats de l'exercice. Enfin, en terminant, le lecteur trouvera une réflexion sur l'impact potentiel d'Internet sur les enquêtes ménages.

## 2. L'ENQUÊTE SUR L'ACCÈS ET L'UTILISATION D'INTERNET AU QUÉBEC

### 2.1 Les objectifs de mesure

L'objectif premier de l'enquête était de déterminer l'appropriation et l'utilisation du réseau Internet par la population québécoise en 1998. À celui-ci, sont venus se greffer d'autres objectifs secondaires : mesurer la perception des non-internautes face à ce nouveau média, mesurer l'intérêt de l'ensemble de la population pour les services gouvernementaux en ligne, etc.

### 2.2 Le contexte méthodologique de l'enquête

Élaborer une base de sondage qui contiendrait tous les internautes québécois est un exercice impossible dans le contexte actuel. Plusieurs facteurs expliquent cet état de fait.

Dans un premier temps, il faut s'interroger sur la caractéristique qui devrait être retenue pour constituer l'unité d'inscription dans une telle liste. Par exemple, dans le cas des ménages, l'unité d'entrée dans la base de sondage peut être l'adresse civique du logement habité pour une enquête en face-à-face ou encore le numéro de téléphone dans le cas d'une enquête téléphonique. Pour les internautes, on pourrait croire que l'adresse de courriel (email) serait un choix approprié. Or, cette hypothèse de travail ne résiste pas à un examen sérieux. D'une part, ce ne sont pas tous les internautes québécois qui ont une adresse de courriel, à preuve, l'enquête a révélé que près de 12 p. cent des personnes qui fréquentent Internet au moins une fois par semaine n'utilisent pas le courrier électronique. D'autre part, il faut faire l'hypothèse que plusieurs internautes partagent la même adresse électronique (adresse familiale), donc qu'ils ne sont pas individuellement dénombrables.

Décider de restreindre l'inférence à la population des internautes québécois qui possèdent une adresse peut sembler une solution, mais encore là de nombreux problèmes se posent. En premier lieu, il faut considérer qu'un bon nombre d'entre eux possèdent plusieurs adresses : une pour le bureau, une à la maison, une pour les communications dans les forums de discussion, etc. Ensuite, s'il est envisageable de contacter l'ensemble des fournisseurs de services Internet du Québec — en 1998, ils sont environ une cinquantaine — et de leur demander la liste des adresses de leurs abonnés, il serait beaucoup plus ardu de contacter tous les employeurs du Québec qui fournissent une adresse de courriel à leurs employés pour leur demander une telle liste. Enfin, plusieurs québécois se branchent à Internet via un fournisseur outre-frontière ou, plus simplement, obtiennent leur adresse de courriel via un serveur de courrier comme hotmail.com, usa.net ou yahoo.com pour ne nommer que ceux-là. Donc, ceux-là, impossible de les retracer tous.

Conclusion : en 1998, il est aberrant de prétendre construire une telle base de sondage en utilisant l'adresse de courriel comme unité d'inscription. Le problème c'est qu'il s'agit de la meilleure, voire de la seule avenue possible. Donc, face à ce problème, on a opté pour une enquête à deux degrés : ménage (1) – individu (2) qui a permis de rejoindre par le téléphone (génération aléatoire de numéros) un certain nombre d'internautes et de les inviter à se rendre sur Internet pour compléter un questionnaire sur leurs habitudes en ligne.

### 2.3 Le questionnaire de l'enquête

Le questionnaire compte cinq sections distinctes : (1) l'introduction qui détermine l'admissibilité du répondant à participer à l'enquête sur le Web, (2) une section concernant les habitudes et perceptions des internautes, dont le profil (utilisation trop restreinte, pas d'utilisation du Web, etc.), (3) une section touchant les perceptions des québécois qui n'ont jamais utilisé Internet, (4) la section des mesures sociodémographiques et enfin, (5) le questionnaire disponible sur le réseau Internet visant les internautes réguliers (branchés au moins une fois par semaine) et utilisateurs du Web. À titre indicatif : dans son ensemble, le questionnaire mesurait plus de 200 variables.

On peut s'interroger sur l'intérêt de mener une telle enquête à la fois par le biais d'entrevues téléphoniques et sur le Web. Il suffit d'avoir fait quelques enquêtes avec comme objectifs de mesurer des caractéristiques technologiques pour comprendre. Malgré toute leur

bonne volonté, les répondants sont souvent incapables de répondre avec précision à de telles questions. Dans le cas de cette enquête, on cherchait à évaluer si les internautes québécois avaient un équipement capable de lire des fichiers PDF, d'écouter des sons (RealAudio, WAV, AIFF, AU), de voir des fichiers d'images animées (MOV), d'exécuter des scripts JAVA, et de visualiser des couleurs d'arrière-plan. La largeur de l'écran de navigation, donnée très importante pour les programmeurs, faisait aussi partie des objectifs techniques de mesure.

### 2.4 La technologie pour la collecte des données sur le Web

Lorsqu'un répondant à l'enquête était identifié au cours de l'entrevue téléphonique, comme étant un utilisateur régulier du réseau Internet et du Web, on l'invitait à se rendre sur le site de l'enquête à l'adresse <http://enquete.bsq.gouv.qc.ca/>. Pour éviter qu'une personne ne tombe par hasard sur le site et complète un questionnaire sans avoir été sélectionnée pour le faire, l'accès au site était protégé par un nom d'utilisateur et un mot de passe. Afin de faciliter la tâche du répondant, le nom d'utilisateur était son numéro de téléphone et le mot de passe, les quatre derniers chiffres de celui-ci. Ainsi, avant même de commencer la collecte des données, le site comptait quelque 8000 usagers, correspondant à tous les numéros de téléphone aléatoires qui composaient l'échantillon.

L'usage du mot de passe, en plus de ne laisser passer que les répondants sélectionnés, permettait de renvoyer à l'endroit approprié, un répondant qui aurait « planté » au moment où il tentait de compléter le questionnaire pour la première fois. Ensuite, une fois le questionnaire complété, le programme de gestion de l'échantillon détruisait l'accès de telle sorte qu'un répondant ne pouvait revenir sur le site et compléter un second questionnaire.

La technologie de programmation du questionnaire permettait les sauts de questions, l'usage de questions filtres et de reposer des questions qui auraient été laissées sans réponse. Il est important de souligner l'apport des spécialistes du RISQ pour la réalisation de cette phase technologique de l'enquête. Ils ont été les premiers au Québec à faire des enquêtes sur le Web et sans leur expertise, le BSQ n'aurait pu procéder à une telle collecte dans des délais semblables.

Cette méthode s'avère une des façons les plus efficaces de récolter de l'information au moyen d'un questionnaire autoadministré, dépassant les possibilités offertes par la technologie DBM (Disk-by-

mail). En effet, l'usage du Web permet de s'affranchir des problèmes causés par les divers systèmes d'exploitation sur le marché. Par exemple, si on désire recueillir des données à l'aide d'un questionnaire autoadministré via le DBM, il faut s'assurer d'en avoir une version qui va fonctionner sous PC (une version DOS pour les systèmes plus anciens et une version Windows 95 / 98 / NT pour les plus modernes), ensuite, une version exécutable sur le système 7 ou 8 d'Apple pour les Macintosh et enfin, pourquoi pas, une version UNIX pour rejoindre les usagers haut de gamme. Le Web, à l'opposé, est conçu pour être accessible, sans discrimination, à tous ces systèmes. On enregistre par conséquent des économies de temps notables au moment de la préparation de l'enquête.

## 2.5 Le suivi des répondants

On identifie un répondant au téléphone, on lui indique l'adresse d'un site Web et on l'invite à s'y rendre pour compléter un questionnaire. Pour faire en sorte que cette procédure n'entraîne qu'un nombre minimal de pertes, des mesures rigoureuses de suivi ont été mises en place. D'abord, le lendemain de l'entrevue téléphonique, on envoyait un message de courrier électronique à tous les répondants qui possédaient une adresse de courriel. Ensuite, au moment de l'entrevue téléphonique, le répondant devait indiquer le moment (jour) où il avait l'intention de se rendre sur le site pour compléter son questionnaire. Dans le cas où la promesse n'était pas tenue, on le relançait dès le lendemain pour obtenir un nouvel engagement de sa part. Si un deuxième rappel s'avérait nécessaire, on offrait au répondant la possibilité de compléter le questionnaire en entrevue téléphonique, en sacrifiant malheureusement la partie technologique du questionnaire.

## 2.6 La confidentialité des réponses

Le BSQ, comme tout organisme statistique, doit s'assurer que les données recueillies dans le cadre d'une enquête demeurent confidentielles. Dans le cadre de l'enquête dont il est question ici, la confidentialité des répondants était assurée en deux temps. D'abord, les caractéristiques socio-démographiques étaient mesurées lors de la phase téléphonique de l'entrevue. Ensuite, les données circulant sur le réseau étaient encryptées avec un logiciel de compression des données dont la clé d'encryptage changeait quotidiennement.

## 2.7 Le résultat de la collecte des données sur le Web

L'enquête a permis d'identifier 484 internautes hebdomadaires, dont 455 utilisaient le Web. Parmi ceux-ci, 340<sup>3</sup> ont été invités à participer à l'enquête sur le Web. Résultat décevant : seulement 173 ont accepté l'invitation qui leur était lancée. Parmi ceux-ci, 119 ont effectivement complété un questionnaire sur Internet, 34 ont été convertis par le téléphone après avoir fait défaut de compléter le questionnaire après deux rappels incitatifs et 20 ont été perdus.

## 2.8 Conclusion de l'expérience

En termes techniques, l'expérience a été un succès. Un questionnaire autoadministré qui se valide au fur et à mesure que les réponses y sont saisies, qui pose uniquement les questions pertinentes et qui fait automatiquement les sauts de section est un outil très performant. Avec un volume suffisant de questionnaire, les coûts de programmation pourraient rapidement être amortis, engendrant ainsi des économies d'échelle intéressantes. Dans le cas présent, le nombre restreint de répondants n'a malheureusement pas eu cet impact.

Par ailleurs, le projet, qui mettait en cause un environnement parfois chaotique (le réseau Internet), n'a pas semblé souffrir de cette particularité. La presque totalité des répondants ont réussi à répondre aux questions technologiques qui mettait à l'épreuve certaines caractéristiques de leur équipement. On peut conclure que, le cas échéant, ceux qui ont « planté » à certains endroits se sont rebranchés sur le site pour compléter leur questionnaire.

D'un côté plus négatif, il faut déplorer l'intérêt un peu faible à vouloir se brancher sur le site pour compléter le questionnaire. Près de la moitié des répondants (167 sur 340) ainsi sollicités ont préféré poursuivre l'entrevue au téléphone. Parmi ceux-ci, certains invoquent le fait que leur accès à Internet est restreint à leur lieu de travail et qu'il ne sont pas nécessairement prêt à sacrifier leur pause-café ou une partie de leur heure de dîner pour prendre part à l'enquête. D'autres mentionnent que le contrat d'accès avec leur fournisseur de connectivité est

<sup>3</sup> Les 115 répondants à qui on n'a **pas proposé** de se rendre sur le Web pour compléter un questionnaire étaient soit : (1) des répondants appartenant à des ménages auprès desquels un refus avait été essuyé, ou (2) des répondants appartenant à des ménages rejoints en toute fin de collecte et qui ne pouvaient bénéficier du délai d'attente normal pour compléter leur questionnaire.

limité à 10 heures par mois et que ce quota est presque atteint. Dans la plupart de ces cas de refus, on peut comprendre que les coûts sous-jacents à la participation (directs dans certains cas et indirects dans d'autres) sont à l'origine de ces refus de continuer l'entrevue sur le Web.

À l'instar de ce qui devra être envisagé pour tenir des entrevues avec des abonnés de la téléphonie cellulaire, on devrait se questionner sur la pertinence éventuelle de « rémunérer » ou de compenser ces répondants pour les frais encourus par le fait de répondre à l'enquête.

### **3. L'IMPACT POTENTIEL DU RÉSEAU INTERNET SUR LES ENQUÊTES FUTURES**

À la lecture du présent article, on a pu constater un impact positif d'Internet sur la réalisation d'enquêtes. Ce nouveau média permet de faire des enquêtes par questionnaires autoadministrés de façon très efficace. L'avenir démontrera sans doute d'autres possibilités de ce média dans le domaine des enquêtes.

Par contre, à brève échéance, il est possible que certaines particularités d'Internet s'avèrent des obstacles à la réalisation d'enquêtes. À cet égard, le cas des enquêtes ménages avec génération aléatoire de numéros téléphone mérite qu'on s'y attarde.

D'abord, si les communications téléphoniques transitent par le réseau Internet<sup>4</sup>, rien n'empêche la transportabilité des numéros de téléphone, de la même façon que les adresses URL sont transportables. En effet, si on inscrit un site Web chez un fournisseur d'accès québécois sous le nom de [www.MonSiteWeb.com](http://www.MonSiteWeb.com), il est tout à fait possible de déménager ce site chez un fournisseur japonais situé à Tokyo tout en gardant la même adresse URL, seule l'adresse IP ayant changé. Ainsi, un québécois pourra potentiellement déménager son numéro de téléphone outre-frontière, et les immigrants pourront apporter des nouveaux numéros de téléphone.

Ensuite, pourquoi se restreindre à des numéros de téléphone du style (000) 000-0000? Il est clair que la possibilité d'utiliser des numéros fantaisistes va exister. Cette alternative sera-t-elle alléchante pour les abonnés de la téléphonie IP ou de son successeur? Ces deux caractéristiques vont potentiellement compliquer la tâche qui consiste à générer un échantillon aléatoire de numéros de téléphone.

En terminant, abordons le point qui se révélera peut-être le plus important. Plusieurs analystes s'entendent pour parler d'une convergence imminente entre les technologies de communication actuelles (Internet, le téléphone, la câblodistribution, etc.). Des mécanismes pour résister aux intrusions accompagneront vraisemblablement cette convergence. Les compagnies de téléphone offrent d'ailleurs déjà la technologie qui permet à leurs abonnés de filtrer leurs appels. Il est probable que les outils disponibles avec le réseau Internet vont accentuer ces possibilités. Face à un nombre sans cesse croissant d'appels venant de compagnies de télémarketing et de représentants de tout acabit, quelle proportion des nord-américains vont se servir des possibilités accrues d'Internet pour contrer ces intrusions, fermant la porte du même coup aux organismes statistiques? Quel impact cette réaction aura-t-elle sur les taux de réponse, les biais dans les estimations et la validité même des enquêtes auprès des ménages?

### **RÉFÉRENCES**

Réseau interordinateurs scientifique québécois. (Page consultée le 31 août 1998). Bienvenue au RISQ [en ligne]. Adresse URL: <http://www.risq.qc.ca/>

Centre francophone d'informatisation des organisations. (Page consultée le 31 août 1998). CEFRIO – Page d'entrée – [en ligne]. Adresse URL: <http://www.cefrio.qc.ca/>

<sup>4</sup> Cette technologie existe d'ailleurs déjà sous le nom de téléphonie sur IP.