



NOTE D'ORIENTATION

Gestion du risque d'indiscipline fiscale:

Progrès accomplis dans la mise au point d'instruments de recherche sur l'Internet à l'intention des administrations fiscales

Préparée par

Forum sur l'administration de l'impôt
Sous-groupe sur la discipline fiscale

Approuvée par

Comité des affaires fiscales
Septembre 2004



SOMMAIRE

PRESENTATION.....	3
<i>Contexte</i>	3
<i>Mise en garde</i>	3
<i>Renseignements complémentaires</i>	3
RESUME.....	4
INTRODUCTION	5
1 EXPERIENCES	8
<i>Organisation</i>	8
<i>Risques</i>	9
<i>Méthodes de recherche</i>	9
<i>Expérience de l'Allemagne et des Pays-Bas</i>	11
<i>Utilisation des données extraites d'Internet</i>	13
2 COMPOSANTES FONCTIONNELLES DES INSTRUMENTS DE RECHERCHE SUR INTERNET.....	14
(1) <i>Surveillance du commerce sur Internet</i>	14
(2) <i>Repérage des sites Web à haut risque</i>	15
(3) <i>Identification du fournisseur de services Internet</i>	16
<i>Collaboration en cours</i>	17
ORIENTATION	19

PRESENTATION

Contexte

En janvier 2003, le Forum sur l'administration de l'impôt a décidé d'entreprendre une étude portant sur l'expérience de différents pays en matière de mise au point d'instruments de recherche sur l'Internet à des fins de contrôle de la discipline fiscale.

A cet effet, un groupe informel a été créé, composé de responsables des services fiscaux des pays suivants, membres de l'OCDE : Autriche, Canada, France, Allemagne, Irlande, Italie, Pays-Bas, Espagne, Suède, Royaume-Uni et États-Unis. Les travaux ont été coordonnés par l'administration des impôts et des douanes des Pays-Bas et menés sous la direction du Sous-groupe sur la discipline fiscale du Forum sur l'administration de l'impôt. Ce groupe a préparé le présent rapport qui fait état des progrès accomplis dans la mise au point d'instruments de recherche sur Internet et fournit des orientations pratiques pour l'utilisation de ces outils.

Mise en garde

Chaque administration fiscale se situe dans un environnement spécifique au sein duquel elle administre son système fiscal. L'environnement politique et législatif de même que les pratiques administratives et les cultures diffèrent d'une juridiction à l'autre. En conséquence, une approche type de l'administration fiscale n'est sans doute ni efficace ni souhaitable dans une situation donnée.

C'est dans cet esprit que doivent être compris et interprétés les documents constituant la série des notes d'information et d'orientation de l'OCDE. Dans l'étude des pratiques adoptées par un pays, il faut toujours veiller à tenir compte de l'ensemble des facteurs qui ont déterminé l'approche considérée.

Renseignements complémentaires

Toute question concernant les différents points évoqués dans la présente note d'orientation doit être adressée à Richard Highfield (Chef, Division du commerce électronique et de l'administration fiscale et des impôts sur la consommation), téléphone : ++33 (0)1 4524 9463 ou courrier électronique (Richard.Highfield@oecd.org).

RESUME

La présente note aborde plusieurs aspects de l'utilisation des instruments de recherche sur Internet à l'intention des administrations fiscales (pour le contrôle de la discipline fiscale), à savoir :

- le contexte général de l'émergence du commerce électronique et les problèmes d'indiscipline fiscale qui s'y rapportent et qui justifient le recours à des instruments de recherche sur Internet ;
- une présentation des expériences de plusieurs pays en matière d'utilisation d'instruments de recherche sur Internet ;
- une description des composantes fonctionnelles des instruments de recherche sur Internet susceptibles d'être utilisés dans le cadre des procédures d'identification des risques par une autorité fiscale.

Compte tenu de la nature importante, mais encore embryonnaire, de ce travail, la présente note donne en conclusion une orientation aux pays membres pour qu'ils poursuivent leurs efforts dans ce domaine, et souligne la nécessité d'une collaboration suivie pour la mise au point et le partage des techniques et des instruments de recherche les plus efficaces ainsi que d'un échange d'expériences à cet égard, en particulier :

- Les autorités fiscales sont encouragées à mettre au point des instruments de recherche sur Internet conformément aux fonctionnalités décrites pour : 1) surveiller le commerce sur Internet et analyser les tendances ; 2) repérer les sites à haut risque ; et 3) identifier les fournisseurs de services à Internet.
- Les autorités fiscales sont encouragées à collaborer pour la mise au point et le partage des techniques et des instruments de recherche sur Internet les plus efficaces ainsi qu'à échanger leurs expériences concernant l'utilisation de ces instruments et techniques.
- Les autorités fiscales sont encouragées à collaborer pour la mise au point de techniques et d'instruments de recherche sur Internet efficaces ainsi qu'à échanger leurs expériences en matière d'utilisation de ces instruments et techniques.

INTRODUCTION

- 1 Le développement des échanges internationaux de services et en particulier du commerce électronique a suscité de nouveaux risques d'indiscipline fiscale. Les risques évoqués dans le présent document concernent les entreprises qui vendent des biens ou des services par Internet, activité dénommée "commerce électronique" au sens large. Les risques d'indiscipline fiscale peuvent être considérés tant à un macro-niveau qu'à un micro-niveau du commerce électronique. Au macro-niveau, il existe des facteurs généraux de risque qui contribuent à l'indiscipline fiscale et il existe aussi des facteurs de risque propres au secteur d'activité. Les facteurs généraux de risque peuvent être rattachés à l'environnement d'Internet et sont, en partie, le résultat de la manière dont Internet s'est développé. De fait, Internet est passé d'un support destiné à des universitaires pour s'échanger des informations textuelles à un support au service des entreprises, et cela grâce aux progrès technologiques qui ont permis des interactions entre utilisateurs, la mise en place de systèmes de paiement électronique et la réalisation complète d'échanges dans le cas de produits livrés par voie numérique¹. Le fait que les pouvoirs publics n'ont pas voulu interférer dans ce développement est un autre facteur important qui est venu s'ajouter à l'évolution d'Internet. Tous ces facteurs ont contribué à créer un environnement qui présente de nouveaux risques pour les administrations fiscales. L'incidence des nouveaux risques d'indiscipline fiscale est largement reconnue, mais l'impact du commerce électronique sur l'importance de cette indiscipline n'est pas connue. On ne dispose à l'heure actuelle d'aucune méthode permettant de mesurer l'importance de l'indiscipline fiscale dans un environnement de commerce électronique.

- 2 En 2003, le Comité des affaires fiscales a présenté des rapports qui signalent plusieurs facteurs généraux de risque liés à l'environnement, notamment :
 - absence d'implantation physique permettant d'identifier la juridiction fiscale dont relève une source de revenu donnée ;
 - les formes traditionnelles de contrôle comme les contrôles physiques aux frontières ne sont pas applicables dans un environnement virtuel où les limites géographiques sont invisibles ;
 - les sites Web peuvent être pilotés à distance et sont faciles à déplacer d'un régime fiscal à l'autre ;
 - la preuve de l'identité est souvent très faible sur Internet et peut être protégée par la législation sur la protection des personnes ;
 - émergence de nouveaux modèles d'entreprise ;
 - désintermédiation ;
 - mécanismes de paiement pouvant fonctionner en dehors des systèmes bancaires existants.

¹ Les systèmes de paiement des plus gros commerçants dans le domaine du commerce électronique sont sécurisés grâce à la technologie SSL ; en conséquence, le développement des serveurs SSL est une bonne indication du développement du commerce électronique. Cf.: http://www.securityspace.com/s_survey/sdata/200402/domain.html

- 3 Le développement des services proposés sur Internet et la complexité de l'environnement numérique se sont soldés par un accroissement du risque d'échapper à un contrôle, notamment avec l'absence de suivi des vérifications, des données incomplètes, l'absence de contrôle du système tel que l'accès et l'authentification, et la situation des livres et des registres bien souvent en dehors de toute juridiction fiscale.
- 4 A un macro-niveau, des instruments de recherche sur Internet pourraient servir à identifier les facteurs généraux qui contribuent à l'indiscipline fiscale en particulier par la surveillance des échanges réalisés par Internet. Cela permettrait de mieux comprendre les facteurs généraux de risque inhérent à un environnement de commerce électronique. A une plus petite échelle, il serait alors possible de tracer le profil de secteurs particuliers qui font du commerce électronique afin de mieux comprendre les risques d'indiscipline fiscale propres à ce secteur. Certains secteurs d'activité peuvent présenter un degré élevé de risques propres du fait de leur modèle d'entreprise et de leur environnement législatif. Les causes de défaut d'enregistrement ou de non-déclaration des revenus imposables sont déjà plus pertinentes si elles sont ventilées par secteur d'activité, mais en outre elles fournissent des informations utiles aux fonctionnaires des impôts dans une optique tant d'élaboration de la politique fiscale que de sa mise en application. Au micro-niveau, le fait de repérer dans un secteur d'activité des entités qui sont en défaut d'enregistrement ou ne déclarent pas des revenus imposables permet de valider des niveaux de risque ainsi que les causes sous-jacentes qui contribuent à l'indiscipline fiscale.
- 5 Certes, il faut comprendre les causes de l'indiscipline fiscale dans un secteur d'activité, mais il faut aussi des informations pour en mesurer l'importance. Les instruments de recherche sur Internet pourraient aider à fournir les informations sur la taille globale du marché, le nombre d'entreprises concernées, ainsi que les informations permettant de mesurer le niveau d'indiscipline fiscale dans un secteur donné. En outre, il faut tenir compte de l'environnement dans lequel se trouvent les entreprises d'un secteur donné pour évaluer correctement les causes sous-jacentes des risques identifiés.
- 6 Certains secteurs présentent des risques élevés du fait de leur environnement législatif et commercial. Les exemples suivants de modèles d'entreprise illustrent bien ce point :
 - enchères et troc : les sites de vente aux enchères et de troc peuvent faire partie de ce qu'on appelle l'économie « grise » ou « souterraine » ;
 - jeux et *divertissement pour adultes* : de nombreux sites sont situés ou censés être situés dans des juridictions de type paradis fiscaux.
- 7 Au micro-niveau, le but est d'identifier les risques concernant des individus ou des petits groupes de contribuables. Les principales préoccupations portent sur l'étude du risque de défaut d'enregistrement auprès d'un régime fiscal ou de minoration des revenus déclarés. Les facteurs sous-jacents à ces risques sont notamment :
 - l'utilisation de prête-nom ou de fausses identités ;
 - la multiplication d'entités écrans dans des juridictions offshore pour masquer la vraie propriété ;
 - les enchères en ligne ;
 - la réapparition du troc sur Internet ; et
 - les problèmes de prix de transfert dans le commerce entre entreprises se traduisant par la circulation des revenus à destination de juridictions à bas taux d'imposition.

- 8 En outre, il existe de nombreuses conventions bilatérales en particulier dans le commerce entre entreprises qui opposent des restrictions à l'accès à des sites, ce qui rend encore plus difficile l'identification des risques.
- 9 En résumé, il est évident que le développement du commerce sur Internet pose de nombreux problèmes aux administrations fiscales. Il est nécessaire d'en savoir plus sur un certain nombre d'aspects généraux du commerce électronique tels que :
 - la nature et la portée de certains types de commerce électronique ;
 - la manière dont le commerce électronique évolue, c'est-à-dire le développement de domaines déjà identifiés (par exemple, le nombre de sites impliqués), la disparition de certains et l'apparition de nouveaux ;
 - l'identification des activités de commerce électronique potentiellement à haut risque ; et
 - le volume des taxes et des impôts sur les transactions de commerce électronique qui ne sont pas payées.
- 10 Une surveillance de l'évolution du commerce sur Internet dans une optique fiscale pourrait révéler des activités économiques émergentes susceptibles de présenter à l'avenir des risques pour les administrations fiscales. A un micro-niveau, il faudrait disposer d'informations sur :
 - les biens et les services fournis par l'intermédiaire de sites individuels ou d'un groupe de sites identifiés ;
 - le résultat des enquêtes de sorte que ces informations puissent servir à évaluer et contrôler les commerçants qui font du commerce électronique.
- 11 Avec la création d'une base de données des sites intéressant les administrations fiscales, celles-ci pourraient disposer de données pour élaborer leurs stratégies de contrôle.

1 EXPERIENCES

- 12 Au cours de la réunion du groupe à Utrecht les 11 et 12 septembre 2003, les pays ont été invités à présenter brièvement leur expérience de l'utilisation d'instruments de recherche. Sur la base des présentations faites par les délégations d'Autriche, du Canada, de France, d'Allemagne, d'Irlande, d'Italie, des Pays-Bas, d'Espagne, de Suède, du Royaume-Uni et des États-Unis, le présent exposé livre les pratiques courantes en matière de recherche sur Internet. Comme cet exposé se fonde sur l'expérience d'un nombre limité de pays, il ne prétend pas brosser un tableau complet des méthodes utilisées par les administrations fiscales des pays membres de l'OCDE. Son but est de fournir aux pays membres des informations sur les approches couramment adoptées dans un certain nombre de pays membres de l'OCDE.

Organisation

- 13 Toutes les administrations fiscales reconnaissent les risques particuliers d'indiscipline liés au commerce sur Internet, en particulier le commerce de services électroniques. Très récemment, beaucoup ont pris des mesures d'ordre organisationnel pour s'attaquer aux nouveaux risques d'indiscipline fiscale. Les méthodes d'organisation adoptées pour surveiller ces risques diffèrent d'un pays à l'autre, mais d'une manière générale, on relève un certain nombre de similitudes.
- 14 La plupart des administrations fiscales représentées dans le groupe ont dépassé le stade où la surveillance et l'analyse du commerce sur Internet sont laissées à l'initiative d'un petit nombre d'experts en informatique employés par chacune d'elles. Dans le cadre de leur stratégie de contrôle général, une attention toute particulière est portée à l'analyse des risques liés au commerce sur Internet. Dans la plupart des pays, des analyses de risques et des programmes de contrôle sont préparés de façon centrale par des départements spécialisés dans le contrôle de la discipline fiscale, et ceux-ci affectent une part croissante de leurs ressources à des projets concernant Internet et au contrôle d'entreprises impliquées dans le commerce électronique.
- 15 A titre d'exemple, la Direction générale des programmes d'observation de l'Agence du revenu du Canada a lancé un certain nombre de projets de "vérification des entreprises sur Internet". Ces projets sont destinés non seulement à mieux comprendre l'environnement technologique et l'évolution des pratiques des entreprises mais aussi à découvrir les domaines potentiellement à haut risque d'indiscipline fiscale.
- 16 Certaines autorités fiscales (Autriche, France, Allemagne, Irlande, Italie et Suède) sont allées plus loin en constituant des équipes spéciales d'enquêteurs chargées de surveiller les risques liés à Internet. Par exemple, en Autriche, un centre de compétences (Kompetenzzentrum Internet und Cybercrime) a été créé pour réunir des informations sur le commerce mondial sur Internet.
- 17 Outre les changements d'organisation, la surveillance et l'analyse du commerce sur Internet d'un point de vue fiscal exigent un personnel présentant des compétences

bien spécifiques. Un certain nombre d'administrations fiscales affectent des ressources à la formation de spécialistes du contrôle informatique (contrôleurs spécialisés) et de conseillers techniques, familiarisés avec l'utilisation d'instruments de recherche sur Internet dans le cadre de procédures de contrôle.

Risques

- 18 Les recherches sur Internet se font pour diverses raisons. D'une part, il s'agit de réunir des informations générales sur les tendances du commerce sur Internet ; d'autre part, une recherche plus spécifique vise à repérer les sites de contribuables indisciplinés, et à extraire des informations directement utilisables pour des vérifications fiscales.
- 19 La majeure partie des informations réunies sur Internet présentent un caractère général et visent principalement à tracer les tendances du commerce sur Internet. La plupart des administrations fiscales portent une grande partie de leur attention à la surveillance du commerce international des marchandises. La croissance du commerce sur Internet a entraîné une forte augmentation du nombre des petites livraisons. Cela présente plusieurs nouveaux risques tels que la non-déclaration ou la minoration de la vraie valeur des produits importés ou l'importation de marchandises illégales. En suivant les tendances du commerce sur Internet, les administrations fiscales cernent mieux les domaines de risque. La recherche d'information générale porte surtout sur certains secteurs économiques qui fonctionnent sur Internet et s'accompagnent de risques supplémentaires d'indiscipline fiscale. A cet égard, on cite généralement le secteur des divertissements pour adultes, les jeux et les sites d'enchères. En enquêtant sur le commerce sur Internet dans certains secteurs, les administrations fiscales espèrent parvenir à cerner l'ampleur probable de l'indiscipline fiscale. A ce jour, les recherches menées par différents pays n'ont guère été concluantes à cet égard.
- 20 La recherche sur Internet sert aussi à obtenir des informations plus spécifiques. Un projet allemand de mise au point d'un instrument de recherche sur Internet a pour objet d'identifier les entreprises qui font du commerce sur Internet et qui ne se sont pas faites enregistrer auprès du fisc ou ne déposent pas de déclaration de revenu. La plupart des autres administrations fiscales essaient d'utiliser les techniques de recherche sur Internet pour repérer les contribuables (indisciplinés), la recherche est alors menée sur la base de profils de contribuables indisciplinés. Certaines recherches sur Internet sont très ciblées. L'Italie, par exemple, a fouillé Internet pour trouver des informations lui permettant de vérifier les revenus déclarés par les représentants juridiques de joueurs de football et de disc-jockeys.
- 21 Aux États-Unis, la technologie des moteurs de recherche a servi à repérer les promoteurs de montages fiscaux frauduleux. Un recoupement des informations trouvées sur Internet avec les données internes des administrations fiscales pourrait aussi apporter une plus grande précision à la classification des secteurs d'activité quant au risque d'indiscipline fiscale à des fins de vérification.

Méthodes de recherche

- 22 Plusieurs méthodes sont utilisées pour recueillir les informations nécessaires afin de traiter différents types de risque. Le choix de la méthode dépend du caractère général ou spécifique des informations recherchées. Pour obtenir des informations générales sur les tendances du commerce sur Internet, les autorités fiscales

s'adressent principalement à des fournisseurs de données spécialisés, tant publics que privés. C'est le cas par exemple des données fournies pour les annuaires d'associations professionnelles ou sectorielles, du matériel de recherche, des statistiques de l'OCDE et d'organismes statistiques nationaux, et de l'abonnement en ligne à divers supports d'informations sur le commerce électronique (Forester, Cyber Atlas).

- 23 Certaines méthodes utilisées pour recueillir des informations spécifiques sont encore très rudimentaires et à forte intensité de main-d'œuvre. Des méthodes plus avancées font appel aux toutes dernières technologies de recherche ; la recherche est alors en grande partie automatisée. Parmi les méthodes les plus rudimentaires mentionnées, on trouve :
- la recherche de publicités et d'indices dans la presse classique et la presse en ligne, qui contiennent des indications sur l'activité économique s'adressant au marché national ;
 - la recherche de sites Web tels que www.mywebsite.com pour les sites hébergés, les sites d'enchères et les forums de discussion ;
 - l'utilisation de moteurs de recherche (Google, Yahoo, Altavista, etc.) pour trouver des noms de domaines ;
 - l'utilisation d'outils de méta-recherche capables d'effectuer des recherches plus spécialisées et personnalisées (par exemple, Copernic).
- 24 Toutes les administrations fiscales ne recueillent pas elles-mêmes les informations. Aux États-Unis, le fisc exploite au mieux l'expertise de ses experts internes en externalisant la procédure de recherche, en utilisant la technologie des moteurs de recherche de tiers sur une base contractuelle. L'externalisation permet à ses experts internes de renforcer leurs compétences avec les progrès en matière de logiciels réalisés dans le secteur concurrentiel du renseignement.
- 25 Pour extraire une sélection de données qui soient gérables tant en taille qu'en contenu, les administrations fiscales appliquent souvent la méthode des profils. Ces profils utilisent principalement une chaîne de mots caractéristiques d'un certain type d'entreprises/secteurs opérant sur Internet. Le résultat de la procédure de recherche peut encore être une base de données contenant des milliers d'adresses de sites. Pour réduire encore le nombre de sites et sélectionner les adresses les plus susceptibles d'être en infraction par rapport au fisc, d'autres techniques sont utilisées. Les informations provenant de sources sur Internet peuvent par exemple être rapprochées des bases de données internes d'une administration fiscale. Les informations internes provenant des déclarations de TVA, de déclarations de revenus pour l'impôt direct et des douanes peuvent être rapprochées d'informations fournies par des tiers extraites de sources d'Internet pour tracer un profil de risque. Bien entendu, ce processus peut se trouver compliqué par le fait que des informations actuelles extraites d'Internet doivent être rapprochées de données internes historiques. Comme les données trouvées sur Internet changent rapidement, il n'est pas toujours possible d'effectuer un rapprochement entre les toutes dernières informations tirées de sites Web et les données internes tirées des propres bases de données de l'administration fiscale.
- 26 La plupart des administrations fiscales fouillent encore manuellement des adresses de sites sélectionnées. Les données élémentaires d'un site Web -- propriété, juridiction, date de création -- sont extraites de la page d'accueil du site, de la page de présentation générale, des détails concernant les contacts, de la base de données « whois » ou du logiciel de localisation². Dans la plupart des pays, il est bien rare

² Pour un exemple de base de données 'whois', voir : <http://www.internic.net/whois.html>.

- qu'un site Web montre le numéro d'identification visible d'un contribuable et il n'a d'ailleurs pas l'obligation de le faire. En conséquence, l'identification des contribuables implique un travail souvent très long à forte intensité de main-d'œuvre.
- 27 Lorsque les résultats des processus initiaux d'identification sont encore non concluants, l'administration fiscale française recourt souvent à une tactique qui consiste à essayer d'identifier les commerçants sur Internet en leur achetant des produits sur Internet. Cette méthode est surtout appliquée dans le cas de sites Web hébergés et de forums de discussion lorsque les numéros des fournisseurs de services sur Internet ne sont pas indiqués ou que le commerce est occulte. Une analyse des messages du courrier électronique fournit des informations sur l'adresse du fournisseur de services Internet, l'identité du propriétaire du site, la banque et des informations sur sa localisation.
- 28 L'Italie a adopté une autre méthode. Dans ce pays, la réglementation de la TVA exige des commerçants sur Internet qu'ils fournissent dès le début de leurs activités des informations sur leur site Web et des données sur leur fournisseur d'accès. L'enregistrement auprès des autorités italiennes est obligatoire pour les entreprises qui se lancent dans une activité de commerce électronique. En outre, les commerçants sur Internet doivent indiquer leur activité sur Internet sur leur déclaration annuelle de TVA.

Expérience de l'Allemagne et des Pays-Bas

- 29 A l'heure actuelle, l'Allemagne et les Pays-Bas mettent au point des instruments de recherche sur Internet plus avancés. En 2002, l'administration fiscale allemande a lancé un projet pour repérer les personnes physiques et morales qui font des affaires sur Internet et ne se sont pas fait connaître du fisc en Allemagne. En 2003, le fisc a commencé à utiliser Xpider, outil spécialement mis au point pour répondre aux besoins de l'administration fiscale allemande. Ses fonctions sont notamment :
- la détection de sites ayant des activités commerciales menées par des entreprises ou des entrepreneurs allemands ;
 - l'identification d'activités de commerce électronique jusque là inconnues qui sont soumises à l'impôt en Allemagne ;
 - l'enregistrement des résultats de façon à pouvoir utiliser la preuve dans la procédure d'évaluation et de contrôle.
- 30 En Allemagne, l'instrument de recherche est capable d'effectuer périodiquement des enregistrements automatiques des entreprises qui fournissent des services sur Internet. L'entreprise susceptible d'être soumise à l'impôt en Allemagne est repérée d'après des catégories spécialisées telles que langue, monnaie, objectif commercial, implantation géographique des bureaux de l'entreprise, adresses des contacts et de l'entreprise, connexion de télécommunications, données relatives au fournisseur d'accès, bandeau publicitaire, etc. Le processus de recherche est encore plus automatisé avec le rapprochement automatique des URL qualifiés et des pages Web avec les données des entreprises qui sont déjà connues du fisc. En outre, l'instrument de recherche peut servir à des objectifs spécifiques tels que des recherches déterminées concernant des domaines d'activité à problèmes comme les « promoteurs » de fraude fiscale et de paradis fiscaux.
- 31 Les Pays-Bas ont lancé un projet à la fin des années 1990, dont le développement est à l'heure actuelle poursuivi en coopération avec le Royaume-Uni. Initialement, le projet portait sur la détection d'activités économiques, avec l'identification des

- fournisseurs et l'extraction des informations pertinentes pour les enquêtes ultérieures. Le recours à la technologie pour détecter, identifier et surveiller le commerce électronique a été étudié en détail, en partie par le biais du développement de prototype. Compte tenu de la nature technique à forte intensité de connaissance du projet, des universités ont aussi été invitées à participer à ce travail. Finalement, les chercheurs ont trouvé qu'il était possible d'obtenir de bons résultats dans la lutte contre le commerce électronique illégal ou frauduleux en utilisant la technologie elle-même pour détecter les entreprises fonctionnant sur Internet et pour identifier les personnes physiques ou morales menant ces activités.
- 32 L'instrument de recherche des Pays-Bas peut aussi servir d'échantillonneur et d'outil d'identification. Utilisé comme échantillonneur, il peut donner une vue d'ensemble quantitative des tendances du commerce sur Internet. A cet effet, il recueille et analyse automatiquement des échantillons représentatifs de sites Web. Il est possible de dégager des tendances en répétant l'échantillonnage à intervalles fixes.
 - 33 Pour analyser Internet, un robot de recherche recueille des informations sur des sites Web. Le résultat est un échantillon représentatif des sites Internet ou un ensemble spécifique de sites répondant à certains critères de sélection. Par exemple, un critère pourrait être que seuls les sites s'adressant à un pays donné peuvent être sélectionnés en prenant, par exemple, la langue des sites comme facteur de sélection. Pour être représentatif, l'échantillon doit comprendre des dizaines de milliers de sites. Toutes les étapes suivies pour produire cet échantillon d'Internet doivent être reproductibles de façon à dégager une preuve et pouvoir vérifier le bon fonctionnement du système.
 - 34 L'échantillon d'Internet est ensuite analysé électroniquement de façon à réaliser un classement des sites qui réponde aux caractéristiques du groupe. Ce classement peut être entièrement automatisé.
 - 35 Pour se prêter à une analyse, il faut que les exemples de sites Web recueillis appartiennent à un groupe d'un intérêt particulier. Le système expert identifie ensuite les caractéristiques distinctives et détermine si un site est pertinent pour l'étude. Le système procède en comparant le contenu des composantes textuelles et HTML des sites à ceux d'autres sites de l'échantillon. L'analyse est un processus itératif, dont les résultats sont évalués par un expert. Ce dernier étudie un certain nombre de sites et sélectionne des exemples positifs et négatifs, c'est-à-dire des sites qui, à son avis, font partie du groupe cible pour l'étude et les sites qui (très probablement) n'en font pas partie. Ces exemples sont de nouveau entrés dans le système afin d'améliorer la qualité des résultats. Cette démarche aboutit en fin de compte à des profils classés de sites Web suspects. Il est également possible d'obtenir des exemples positifs et négatifs de sites à partir d'autres sources tels que les signaux de fonctionnement normal.
 - 36 Lorsqu'il est utilisé comme outil d'identification, un profil des sites Web pertinents est créé. Ce profil est traduit en une ou plusieurs demandes de recherche sur Internet. Les données potentielles identifiantes sont recueillies grâce à la technologie linguistique et aux techniques d'extraction.
 - 37 Le système localise et extrait des données caractéristiques sur les sites Web concernés tels que les codes postaux, les numéros de téléphone et les numéros de TVA. Ces données sont sauvegardées dans un format qui leur permet d'être utilisées ultérieurement à des fins d'identification. En outre, le système extrait les données d'enregistrement du domaine du site Web et les ajoute aux autres caractéristiques. Au cours du processus d'identification, les composantes suivantes sont évaluées :

- le contenu des pages constituant le site Web ;
 - les méta-informations sur le site Web ;
 - les informations “whois” concernant l'enregistreur dans le domaine pertinent.
- 38 Pour optimiser la performance, les pages qui sont les plus susceptibles de contenir les données d'identification sont étudiées en premier. Le processus s'arrête quand des caractéristiques suffisantes ont été trouvées qui permettent d'identifier la personne physique ou morale responsable du site Web. On tente ensuite d'identifier électroniquement le fournisseur d'accès en recherchant les mêmes caractéristiques dans les ensembles de données pertinents. La propre base de données de l'administration fiscale constitue une source utile d'information, mais beaucoup d'autres bases de données extérieures sont également consultées. Si le fournisseur ne peut être identifié électroniquement, les informations du site sont interprétées par un expert et une enquête supplémentaire est lancée.
- 39 Les instruments de recherche allemand et néerlandais utilisent la technologie des robots de recherche. Le robot de recherche commence par plusieurs bons exemples de pages Web intéressantes pour l'administration fiscale. La relation contextuelle entre les mots figurant dans ces pages est déterminée par des algorithmes qui orientent la recherche sur des pages similaires. Le but est de réduire le grand nombre de pages Web sélectionnées qui auraient été trouvées par un moteur de recherche standard. Les instruments de recherche allemand et néerlandais sont capables de trouver des indices de déclaration minorée mais pas de mesurer l'ampleur de la minoration.
- 40 Dans ces deux pays, le recours à une technologie avancée de recherche en est encore à ses balbutiements et n'est pas pleinement intégré à la procédure de gestion des risques des administrations fiscales. Toutefois, ces dernières ont acquis un peu d'expérience et les premiers résultats préliminaires montrent que leurs instruments de recherche sur Internet peuvent être très efficaces dans la détection de situations d'indiscipline fiscale potentielles.

Utilisation des données extraites d'Internet

- 41 Les instruments de recherche sur Internet fournissent aux administrations fiscales des informations précieuses sur les activités des opérateurs économiques. Selon le système de gestion et d'évaluation des risques de l'administration fiscale, les informations recueillies grâce à ces instruments de recherche sont traitées de façon centrale ou transmises à des agences fiscales locales. Dans ce dernier cas, des bureaux décentralisés sélectionnent les cas présentant le plus fort potentiel d'indiscipline fiscale et procèdent à un examen des activités sur Internet et des recettes générées.
- 42 Les renseignements généraux obtenus par secteur d'activité sont aussi utiles pour brosser un tableau de l'environnement technologique, des pratiques des entreprises, afin d'améliorer le matériel de formation, de partager les pratiques exemplaires, et d'élaborer et de décrire des stratégies de contrôle satisfaisantes.

2 COMPOSANTES FONCTIONNELLES DES INSTRUMENTS DE RECHERCHE SUR INTERNET

- 43 De plus en plus, les administrations fiscales reconnaissent les risques particuliers d'indiscipline fiscale liés aux échanges transfrontières sur Internet, risques qui augmentent d'autant plus que le commerce sur Internet se développe de façon rapide et régulière. Pour faire face à ces risques, de nombreuses administrations fiscales ont pris des mesures d'ordre organisationnel qui leur permettent de porter une attention particulière à l'étude et la surveillance du commerce sur Internet.
- 44 Les informations recueillies sur Internet sont en partie de nature générique, et en partie spécifiques. Les autorités fiscales recourent fréquemment à la technologie de base de recherche sur Internet -- avec des moteurs de recherche tels que Google et Yahoo -- pour recueillir des informations sur le commerce sur Internet. Souvent, les résultats de la recherche sont examinés manuellement. Il serait préférable d'utiliser la technologie plus avancée qui met en œuvre la toute dernière technologie de recherche automatisant en grande partie le processus. Dans le fonctionnement de ces instruments de recherche sur Internet, on distingue trois grandes composantes fonctionnelles : (1) la surveillance du commerce sur Internet et l'analyse des tendances, (2) le repérage des sites Web à haut risque et (3) l'identification des fournisseurs de services Internet.

(1) Surveillance du commerce sur Internet

- 45 Comme cela a déjà été dit, l'expérience de plusieurs pays montre que de nombreuses administrations fiscales s'intéressent à l'heure actuelle davantage à la surveillance des échanges internationaux de biens et de services. La recherche d'informations générales est principalement axée sur certains secteurs économiques qui opèrent sur Internet et présentent des risques supplémentaires d'indiscipline fiscale. Pour recueillir des informations générales sur les tendances du commerce sur Internet, la plupart des organismes s'adressent à des fournisseurs de données spécialisées, tant publics que privés.
- 46 Il est impératif de mieux comprendre les schémas du commerce sur Internet, en particulier les différents types d'activités sur Internet et leur ampleur. L'une des premières questions est de savoir comment identifier certains types d'activité de commerce électronique potentiellement à haut risque. Il faut bien reconnaître les caractéristiques distinctives de chaque type d'activité. En outre, il faut surveiller l'évolution des différents types d'activité, de façon à pouvoir prédire quels types vont se développer et quels types vont disparaître.
- 47 Internet est un objet de recherche complexe avec des millions de sites pouvant constituer chacun une cible de recherche. En conséquence, il faut d'abord acquérir une compréhension représentative, complète et structurée d'Internet, avant de chercher à identifier des groupes cibles spécifiques au sein de la population totale

des sites Web. Les instruments de recherche sur Internet peuvent jouer un rôle important dans l'acquisition de cette compréhension, grâce à leur fonction de surveillance.

- 48 Pour parvenir à cette compréhension, il faut analyser le contenu des sites Internet. Compte tenu de la taille énorme et de la nature non structurée d'Internet et aussi de la vitesse à laquelle se font les changements, il n'est pas possible de prendre Internet tout entier comme point de départ pour ce type d'analyse. Il vaut mieux constituer un échantillon de sites Internet. Ce sous-ensemble reste néanmoins très grand en termes absolus car il doit être suffisamment représentatif pour révéler les tendances et permettre des analyses pour identifier les caractéristiques de certains types d'activités sur Internet.
- 49 Il faut analyser l'échantillon pour identifier les activités sur Internet qui sont intéressantes. L'analyse est limitée à un sous-ensemble de sites Web présentant des caractéristiques communes et constituant un groupe cible. Il est possible de tracer le profil des caractéristiques d'un groupe cible par secteur d'activité (voyages, divertissements pour adultes, jeux, produits pharmaceutiques, commerce illégal), par implantation géographique (local, régional, national, international), par thème (drogue, arme) ou tout autre critère pertinent.
- 50 Outre une compréhension du contenu actuel d'Internet, il est nécessaire de se faire une idée de l'évolution qui est en cours. Un moyen puissant d'avoir cette idée est de comparer les résultats de ces analyses d'Internet au fil du temps. On peut procéder de deux manières. Premièrement, on peut tracer des profils sur une période relativement longue en reprenant le même échantillon d'Internet. Ces profils changent au fil du temps au fur et à mesure que le contenu des pages des sites change. Il est alors possible de dégager des tendances pour les sites existants. Deuxièmement, on peut comparer plusieurs échantillons de sites Internet entre eux au fil des mois et des années. Les profils tracés à partir des échantillons peuvent aussi servir à ces comparaisons. En surveillant l'évolution des profils, les fonctionnaires des impôts peuvent percevoir le développement d'activités sur Internet telles que la création de nouvelles activités et l'évolution d'activités existantes.
- 51 Cette analyse de tendance sert non seulement à comprendre l'évolution à un macro-niveau mais aussi à surveiller le comportement des fournisseurs de sites Web. Il peut aussi être intéressant de voir les changements qui interviennent dans la fourniture de biens et de services par des commerçants individuels sur Internet. On peut suivre les relevés du type de commerce électronique mené par un commerçant, si ce commerce se fait par des sites différents, et l'importance du commerce au fil du temps, etc.
- 52 L'analyse de tendance est reconnue pour être un outil important à la disposition des administrations fiscales pour surveiller l'évolution du commerce sur Internet. Les administrations fiscales sont encouragées à mettre au point des logiciels opérationnels pour mieux comprendre le commerce sur Internet, et ainsi améliorer leur capacité à réagir rapidement à des situations à haut risque d'indiscipline fiscale, en apportant des réponses tant législatives qu'administratives.

(2) Repérage des sites Web à haut risque

- 53 L'expérience de certains pays montre que les méthodes utilisées par certaines administrations fiscales pour recueillir des informations spécifiques sont encore très rudimentaires et à forte intensité de main-d'oeuvre. Des méthodes plus

avancées, utilisées par un nombre relativement restreint d'administrations, font appel à une technologie plus avancée avec laquelle le processus de recherche est en grande partie automatisée.

- 54 Si l'on veut examiner un groupe de sites sur Internet, il faut alors identifier tous les sites Web pertinents. Dans le domaine des activités de commerce électronique, il est possible d'affiner la sélection, par exemple, en ne retenant que le commerce électronique ciblé sur un pays donné. Les enquêtes peuvent se concentrer sur des types d'activité préalablement identifiés au cours du processus de surveillance. Manifestement, les moteurs de recherche classiques ne conviennent pas à ces cas de recherche complexes. En effet, avec ces moteurs, c'est l'utilisateur qui doit définir lui-même l'objet de la recherche. Avec ces instruments, l'exhaustivité, la représentativité et la qualité de la recherche restent sujets à caution.
- 55 L'expérience et l'appréciation professionnelles sont également de bonnes sources de préparation manuelle des profils destinés à être utilisés dans le processus de recherche. Il s'agit dans chaque cas de trouver tous les sites existant sur une partie d'Internet qui correspondent plus ou moins à un profil de risque. Ici, « une partie de » signifie que le processus de recherche peut être limité à un domaine d'intérêt particulier, par exemple, un pays, tandis que « plus ou moins » se rapporte au fait que certains sites Web trouvés peuvent ne pas correspondre pleinement au profil mais partager avec lui plusieurs caractéristiques principales, sinon la plupart d'entre elles. Il convient donc de fixer dans le processus de recherche un paramètre correspondant à la marge de tolérance admise dans la correspondance entre les sites et le profil défini.
- 56 Les sites trouvés doivent être classés en fonction de la mesure dans laquelle ils correspondent aux caractéristiques du profil de référence. Lorsqu'un nombre suffisant de sites a été trouvé, la recherche peut s'arrêter. On obtient ainsi un groupe de sites qui correspond assez étroitement au profil. A l'issue de ce processus de recherche, on a d'une part les sites trouvés et, d'autre part, le nombre réel de sites, ce qui peut fournir des informations utiles sur l'ampleur du risque en question. A cet égard, les changements de quantités de biens et de services proposés sur un site Web peuvent être pertinents : y a-t-il des pics ponctuels ou bien l'offre est-elle stable ou en augmentation régulière ?
- 57 Les administrations fiscales sont encouragées à mettre au point des logiciels opérationnels pour repérer les sites Web à haut risque d'une manière présentant un bon rapport coût-efficacité.

(3) Identification du fournisseur de services Internet

- 58 L'expérience de plusieurs pays montre que de nombreuses administrations fiscales examinent encore les adresses de sites Web sélectionnés manuellement.
- 59 Si l'on décide de vérifier, par exemple, si un site Web donné qui fournit des biens ou des services sur Internet est légal et paie ses impôts et taxes en totalité, il faut bien identifier le contribuable³. En outre, il pourrait être utile d'examiner les activités de commerce électronique de chaque fournisseur identifié afin de

³ Il est fait référence ici au document sur "Business Identification" publié dans la série de l'OCDE Orientations pour les administrations fiscales, dans laquelle la question de l'identification est analysée en détail ; Cf. <http://www.oecd.org/dataoecd/4/56/14990201.pdf>

déterminer quelles sont les activités de commerce électronique qui sont menées au fil du temps. On pourrait alors détecter des activités illégales ou frauduleuses sur Internet qui autrement passeraient inaperçues compte tenu des fréquents changements de produits et de services fournis et de sites utilisés ou d'autres variations encore.

- 60 Dans la plupart des cas, il est possible d'identifier le fournisseur à partir des informations fournies sur un site Web. Théoriquement, il devrait y avoir un moyen de localiser et d'extraire les données caractéristiques des sites Web concernés telles que codes postaux, numéros de téléphone et numéros de TVA. Ces données sont sauvegardées dans un format qui permet leur utilisation ultérieure à des fins d'identification. De plus, il faudrait que le système permette d'extraire les données d'enregistrement du domaine du site Web pour les ajouter aux autres caractéristiques. Dans le processus d'identification, il est possible d'évaluer les composantes suivantes :
- le contenu des pages constituant le site Web ;
 - les méta-informations sur le site Web ;
 - les informations de type "whois" relatives à l'enregistreur du domaine pertinent.
- 61 Pour optimiser la performance, les pages qui sont les plus susceptibles de contenir les données d'identification sont étudiées en premier. Le processus s'arrête quand des caractéristiques suffisantes ont été trouvées qui permettent d'identifier la personne physique ou morale responsable du site Web. On tente ensuite d'identifier électroniquement le fournisseur d'accès en recherchant les mêmes caractéristiques dans les ensembles de données pertinents. La propre base de données de l'administration fiscale constitue une source utile d'information, mais beaucoup d'autres bases de données extérieures sont également consultées. Si le fournisseur ne peut être identifié électroniquement, les informations du site sont interprétées par un expert et une enquête supplémentaire est lancée.
- 62 Les administrations fiscales sont encouragées à mettre au point un logiciel opérationnel avancé présentant les fonctionnalités décrites comme constituant un moyen efficace d'identifier les fournisseurs de services Internet.

Collaboration en cours

- 63 L'expérience de plusieurs pays montre que de nombreuses administrations fiscales ont conscience de la nécessité d'enrayer les risques d'indiscipline fiscale résultant d'un environnement de commerce électronique. La collaboration dans la mise au point et le partage des techniques et des instruments de recherche sur Internet les plus efficaces présente un intérêt considérable et devrait profiter à toutes les administrations fiscales participantes. A cet égard, il est utile de savoir que certains pays mettent actuellement au point ou ont mis au point des instruments avancés de recherche sur Internet. En raison des caractéristiques du commerce sur Internet, les informations recueillies à un macro-niveau par une administration fiscale peuvent intéresser d'autres administrations fiscales. Par le biais d'une étroite collaboration, celles-ci pourraient utilement tirer parti des réussites et des échecs des unes et des autres.
- 64 L'utilisation potentielle d'instruments avancés de recherche sur Internet pourrait aussi être utile à d'autres services d'enquête publique tels que la police et les douanes.

- 65 La mise au point et l'utilisation d'instruments de recherche sur Internet en vue de trouver des informations pertinentes et à jour pourrait contribuer à réduire les coûts administratifs et les délais, le travail étant effectué par des ordinateurs en continu.
- 66 Les autorités fiscales sont encouragées à collaborer pour la mise au point de techniques et d'instruments de recherche sur Internet efficaces ainsi qu'à échanger leurs expériences en matière d'utilisation de ces instruments et techniques.

ORIENTATION

- Les autorités fiscales sont encouragées à mettre au point des instruments de recherche sur Internet conformément aux fonctionnalités décrites pour : 1) surveiller le commerce sur Internet et analyser les tendances ; 2) repérer les sites à haut risque ; et 3) identifier les fournisseurs de services à Internet.
- Les autorités fiscales sont encouragées à collaborer pour la mise au point et le partage des techniques et des instruments de recherche sur Internet les plus efficaces ainsi qu'à échanger leurs expériences concernant l'utilisation de ces instruments et techniques.
- Les autorités fiscales sont encouragées à collaborer pour la mise au point de techniques et d'instruments de recherche sur Internet efficaces ainsi qu'à échanger leurs expériences en matière d'utilisation de ces instruments et techniques.