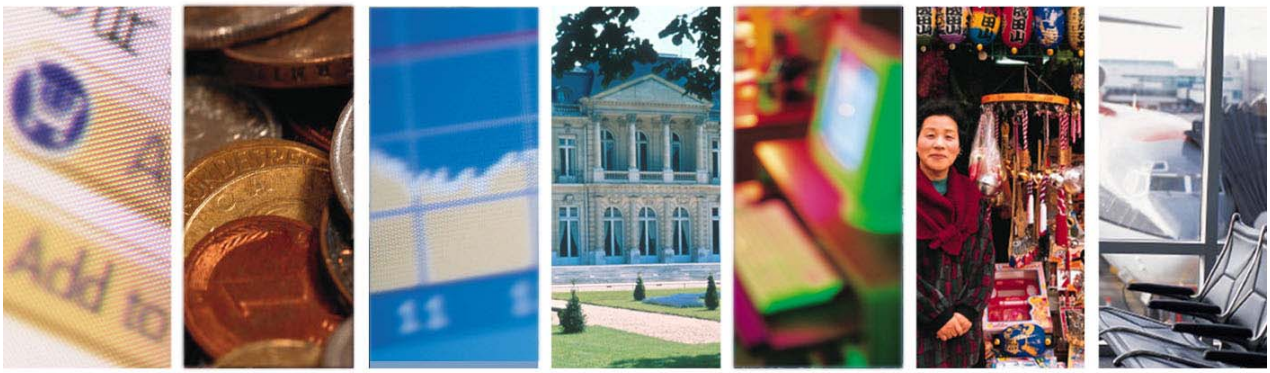
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT



GUIDANCE NOTE

Compliance Risk Management:

Progress with the Development of Internet Search Tools for Tax Administration

Prepared by

Forum on Tax Administration
Compliance Sub-group

Approved by

Committee on Fiscal Affairs
October 2004

CTPA

CENTRE FOR TAX POLICY AND ADMINISTRATION

TABLE OF CONTENTS

ABOUT THIS DOCUMENT 3

Background *3*
Caveat *3*
Inquiries and further information *3*

SUMMARY 4

INTRODUCTION 5

1 EXPERIENCES 8

Organisational issues *8*
Risk areas *9*
Searching methods *9*
Experience of Germany and the Netherlands *11*
Application of the Internet data *13*

2 FUNCTIONAL COMPONENTS OF INTERNET SEARCH TOOLS 14

(1) Monitoring Internet trade *14*
(2) Collection of high-risk websites *15*
(3) Identification of Internet supplier *16*
Collaboration in development *17*

GUIDANCE 18

ABOUT THIS DOCUMENT

Background

In January 2003, the Forum on Tax Administration (FTA) decided to undertake a study of country experiences with the development of Internet search tools for compliance checking purposes.

To carry out this task, an informal group was established, consisting of officials from revenue authorities in the following OECD member countries: Austria, Canada, France, Germany, Ireland, Italy, Netherlands, Spain, Sweden, United Kingdom, and the United States. The work was co-ordinated by the Netherlands Tax and Customs Administration and conducted under the auspices of the FTA Compliance Sub-group. The group has prepared this report to record progress with the development of Internet search tools and to provide some practical guidance in the use of search tools.

Caveat

Each revenue authority faces a varied environment within which they administer their taxation system. Jurisdictions differ in respect of their policy and legislative environment and their administrative practices and culture. As such, a standard approach to tax administration may be neither practical nor desirable in a particular instance.

The documents forming the OECD tax guidance series need to be interpreted with this in mind. Care should always be taken when considering a country's practices to fully appreciate the complex factors that have shaped a particular approach.

Inquiries and further information

Inquiries concerning any matters raised in this guidance note should be directed to Richard Highfield (Head, CTPA Tax Administration and Consumption Taxes Division), phone ++33 (0)1 4524 9463 or e-mail (Richard.Highfield@oecd.org).

SUMMARY

This note focuses on several aspects of the use of Internet search tools for tax administration (i.e. tax compliance checking) purposes, namely:

- the general background of the rise of e-commerce and the corresponding compliance issues against which the need for Internet search tools is being given substance;
- an overview of country experiences with the use of Internet search tools;
- a description of the functional components of Internet search tools that could be applied in the risk identification processes of a revenue authority.

Given the important but embryonic nature of this work, the note concludes with guidance for member countries to continue with their work in this area, including the need for ongoing collaboration in the development and sharing of the most efficient search tools and techniques and to exchange experiences in their use. Specifically:

- Revenue authorities are encouraged to develop internet search tools in accordance with the functionalities described to: 1) monitor the Internet trade and analyse the trends; 2) collect high-risk websites; and 3) identify Internet suppliers.
- Revenue authorities are encouraged to collaborate in the development and sharing of the most efficient Internet search tools and techniques and to exchange experiences in the use of these tools and techniques.
- Revenue authorities are encouraged to collaborate in the development of efficient Internet search tools and techniques and to exchange experiences in the use of these tools and techniques.

INTRODUCTION

1 The growth of trade in international services and especially the growth of e-commerce have resulted in new compliance risks for revenue authorities. The compliance risks referred to in this paper are related to businesses selling goods or services over the Internet, trade that can broadly be termed 'e-commerce'. The compliance risks can be viewed at both the macro and micro levels of electronic commerce. On the macro level, there are general risk factors that contribute to non-compliance and there are also industry specific risk factors. General risk factors can be associated with the Internet environment and are, in part, a result of the manner in which the Internet developed. The Internet evolved from a medium for academia to share textual information to that of a business-driven medium made possible by technological advances allowing user interaction, electronic payment systems, and fulfilment in the case of digitally delivered products[1]. An important factor that added to the Internet's evolution has been governments' reluctance to interfere in its development. Those evolutionally factors have spawned an environment with new risks to tax officials. Although the incidence of new compliance risks is widely acknowledged, the impact of e-commerce on compliance levels is unknown. There are currently no techniques available to measure the level of compliance in an e-commerce environment.

2 Some of the general environmental risk factors that were identified in the 2003 reports to the CFA include the following:

- there is no physical location to identify the tax jurisdiction for a particular source of income;
- traditional forms of control like physical border controls are not applicable in a virtual setting where geographical boundaries are invisible;
- websites can be controlled remotely and may be easily moved across tax regimes;
- proof of identity is often very weak on the Internet and may be protected by privacy laws;
- the emergence of new business models;
- disintermediation;
- payment mechanisms may operate outside existing banking systems.

3 Growth in the services offered via the Internet and the complexity of the digital environment has resulted in increased control risks, including a lack of audit trails, incompleteness of records, a lack of system controls such as access and authentication, and the location of books and records often outside tax jurisdictions.

4 On a macro level, Internet search tools might be used to identify the overall factors that contribute to non-compliance especially through the monitoring of internet

[1] Payment systems of most serious e-commerce traders are secured through the use of ssl-technology; hence the growth of ssl-servers gives a useful proxy of the growth of e-commerce. See: http://www.securityspace.com/s_survey/sdata/200402/domain.html

trade from a compliance perspective. It could assist in understanding the general risk factors present in an e-commerce environment. On a less broad scope, profiles could be made of specific industries that engage in e-commerce to provide a better understanding of specific compliance risks within that industry. Certain industries may have a high degree of inherent risk by virtue of their respective business model and legal environment. The causes of non-registration or of not reporting taxable events are not only more relevant when broken down by industry, but also, provide meaningful information to tax administrators from both a policy and enforcement perspective. At the micro level, identifying specific entities within an industry that do not register or do not report taxable events will validate levels of risk and the underlying causes that contribute to non-compliance.

5 In addition to understanding the causes of non-compliance within an industry, information is also needed to gauge its significance. Internet search tools could help to generate information about the total market size, number of businesses involved, as well as information to gauge the level of non-compliance within an industry. Also, knowledge of the business environment for the industry is needed to properly assess the underlying causes of risks identified.

6 Certain industries pose high risks by virtue of their legal and trading environment. The following examples of business models are illustrative:

- auctions and bartering: sales on auction and bartering websites can be part of what is known as the 'shadow' or 'underground' economy;
- gambling *and adult entertainment:* many websites are located or are supposed to be located in tax haven jurisdictions.

7 At the micro level, there are concerns with the detection of risks for individual or small groups of taxpayers. Detecting the risk of non-registration with a tax regime or the underreporting of tax obligations are the main factors of concern. Factors underlying those risks include:

- the use of nominees or false identities;
- tiered ownership in offshore jurisdictions involving multiple entities designed to conceal ownership;
- online auctions;
- the resurgence of Internet bartering; and
- transfer pricing issues in business-to-business trade involving shifts of income to low tax jurisdictions.

8 Additionally, there are many bilateral arrangements in existence especially in business-to-business trade that set restrictions on the access of websites, thus making the identification of risks more difficult.

9 In summary, it is evident that the development of Internet trade raises many questions for tax administrations. There is a need for insight into a number of general aspects of e-commerce such as:

- the nature and scale of certain types of e-commerce;
- the way e-commerce is evolving, i.e. the development of areas already being identified (e.g. the number of websites involved), the disappearance of areas and the appearance of new ones;
- the identification of potentially high-risk e-commerce activities; and
- the volume of taxes and duties on e-commerce transactions that are not being remitted.

10 Monitoring the developments of trade over the Internet which is tax relevant could render signals of emerging economic activities that might pose future tax revenue risks. On a micro level, information should be available on:

- the goods and services supplied through individual websites or a group of identified websites;
- the outcome of investigations so that documentation could be used in the assessment and auditing of e-commerce traders.

11 Through the creation of a database of tax relevant websites input could be gathered for the control strategies of tax administrations.

1 EXPERIENCES

12 During the meeting of the group in Utrecht on 11 and 12 September 2003 countries were asked to present an outline of the national experiences with the use of search tools. Based on the presentations made by the delegations of Austria, Canada, France, Germany, Ireland, Italy, the Netherlands, Spain, Sweden, the United Kingdom, and the United States an overview is given of the current practices in searching the Internet. As this overview is based on the experiences of a limited number of countries, it does not pretend to give an exhaustive summary of the methods used by revenue authorities in OECD member countries. Rather, it is provided to inform members of the approaches currently used in a number of the OECD member states.

Organisational issues

13 All revenue authorities recognize the special compliance risks that are related to the trade over the Internet, especially the trade in electronic services. Quite recently, many have taken organizational steps to deal with the new compliance risks. Although the organisational methods adopted to monitor these risks diverge between different countries, on a general level there are a number of similarities which can be noted.

14 Most revenue authorities represented on the group have passed the stage where monitoring and investigation of Internet trade are left to the initiative of a few computer experts within each revenue authority. As part of their general control strategy, special attention is being devoted to the analyses of risks that are linked to the Internet trade. In most countries, risk analyses and audit programmes are centrally prepared by specialized compliance/control divisions and these divisions are allocating an increasing amount of their resources to projects which are linked to the Internet and the control of companies involved with e-commerce trade.

15 As an example, the Compliance Programs Branch of the Canada Revenue Agency (CRA) has undertaken a number of “Internet Business Audit” projects. These projects are meant to not only gain a better understanding of the technological environment and evolving business practices but also to uncover potential high risk areas of non-compliance.

16 Some revenue authorities (e.g. Austria, France, Germany, Ireland, Italy, and Sweden) have gone a step further by establishing special investigating teams which are dedicated to monitoring the risks that are related to the Internet. For example, in Austria, a competence centre (Kompetenzzentrum Internet und Cybercrime) has been set up to gather information on global Internet trade.

17 In addition to organizational changes, monitoring and investigating Internet trade from a tax compliance perspective requires staff with special competences. A number of revenue authorities allocate resources for the training of computer audit specialists (EDP auditors) and technical advisors, who are knowledgeable when it comes to the use of Internet (search) tools within the audit process.

Risk areas

18 Investigating the Internet is done for a variety of reasons. On the one hand, the web is searched in order to gather general information about trends in Internet trade; on the other hand, more case-specific research is carried out aimed at identifying websites of non-compliant tax payers, retrieving information that is directly applicable for taxpayer audits.

19 A considerable part of the information gathered on the Internet has a general character and is mainly aimed at discovering trends in the Internet trade. Most tax administrations allocate part of their attention to the monitoring of the international trade of goods. The growth of the Internet trade has caused an enormous increase in the number of small consignments. This poses several new risks such as the non-reporting or underreporting of the true value of the imported good or the import of illegal goods. By monitoring the trends in Internet trade, revenue authorities get a better view of the risk areas. The search for general information is focused mostly on certain economic sectors that operate over the Internet and pose extra compliance risks. Frequently mentioned are the adult entertainment industry, gambling, and auction sites. By investigating the Internet trade of certain sectors, revenue authorities hope to gain some insights as to the likely extent of non-compliance. To date, country research has generally been inconclusive in this respect.

20 Web-searching is also used to get more case specific information. The mission of a German project for the development of a web-search tool is to identify Internet trading businesses that are not registered for taxation or are not filing tax returns. Most other revenue authorities try to use web search technology to identify (non-compliant) taxpayers, whereby the search process is carried out on the basis of profiles of non-compliant taxpayers. Some of the Internet investigations are quite specific. Italy, for example, searched the Internet to gather information to test the income reported by legal representatives of football players and disc jockeys.

21 In the United States, search engine technology has been useful in identifying promoters of abusive tax planning schemes. Combining Internet information with the internal data of revenue authorities could also provide more accuracy in the risk classification of industries for compliance checking purposes.

Searching methods

22 There are several methods being used to collect the information needed for handling different risk areas. The choice of a method depends on whether the information has a general or case-specific character. For the collection of general information about trends in Internet trading revenue authorities mostly turn to specialised data providers, both governmental and commercial. Examples are the data provided in directories of industry associations, research materials, statistics from OECD and national statistical bodies, and online subscription to various news media on e-commerce (Forester, Cyber Atlas).

23 Some of the methods used to collect case-specific information are still quite basic and labour-intensive. More advanced methods make use of the latest searching technology whereby the search process is largely automated. Among the more basic methods mentioned are:

- the searching of traditional and on-line newspapers for advertisements and links that contains indications about economic activities that are directed towards the home market;

- the searching of websites like www.mywebsite.com for hosted websites, auction sites and newsgroups;
- the use of web-based search engines (Google, Yahoo, Altavista, etc.) to find domain names;
- the use of client based meta-search tools which are capable of more specialized, personalized web searches (e.g. Copernic).

24 Not all revenue authorities collect the information themselves. The Internal Revenue Service in the United States leverages the expertise of its internal experts by outsourcing the search process, using search engine technology of third parties on a contractual basis. Outsourcing allows its internal experts to leverage their skills with the skills and software advancements that have taken place in the competitive intelligence industry.

25 In order to retrieve a selection of data that is manageable in size as well as content, revenue authorities frequently apply profiles. These profiles mostly use a string of words that is typical for a certain type of company/branch operating over the Internet. The outcome of the search process might still be a database with thousands of website addresses. In order to further downsize the number of websites and select the addresses that are most likely to be non-compliant, other techniques are used. The information from sources on the Internet may, for instance, be matched with a revenue authority's own internal databases. Internal information from VAT tax returns, direct tax returns, and customs can be matched with third party information from Internet sources to generate a risk profile. Of course, this process may be complicated by the fact that actual information retrieved from the Internet has to be matched with historical internal data. As the data found on the Internet are changing rapidly, it will not always be possible to match the recently captured information from websites with the internal data from a revenue authority's own databases.

26 Most revenue authorities still investigate selected website addresses manually. The identification of the elementary data of a website - ownership, jurisdiction, date of creation – are recovered through the use of websites' home page, the 'about us'-page, contact details, the 'whois'-database or tracking software[2]. In most countries, websites a rarely show a visible taxpayer identification number and there is no requirement for them to do so; as a consequence, the identification of taxpayers is a labour intensive and often time-consuming process.

27 Where the outcome of initial identification processes is still inconclusive, the French tax administration sometimes employs the tactic of trying to identify Internet-traders by making purchases over the Internet. This method is especially applied in cases of hosted websites and newsgroups where ISP-numbers are not available or trade is hidden. An analysis of the e-mail messages provides information about ISP-address, site-owner ID, bank, and location information.

28 Another approach has been taken by Italy where the VAT regulation requires starting Internet-traders to provide information about their website and provider data. Registration with the Italian Registration Authority is compulsory for companies performing e-commerce activities. In addition, Internet-traders have to indicate their Internet-activity on their yearly VAT returns.

[2] For an example of a 'whois'-database see: http://www.internic.net/whois.html.

Experience of Germany and the Netherlands

29 Germany and the Netherlands are developing more advanced Internet search tools. In 2002, the German tax authorities have set up a project to detect natural and legal persons that are conducting business over the Internet not known to the German tax authorities. In 2003, they started using Xpider, a tool especially developed to meet the needs of the German tax authorities. Its functionality includes:

- the detection of sites with commercial activities conducted by German entrepreneurs or businesses;
- the identification of previously unknown e-commerce activities that are subject to tax in Germany;
- the recording of results so that evidence can be used in the process of assessment and auditing.

30 In Germany, the search tool is capable of periodically making automated records of businesses supplying services on the Internet The business that might be liable for tax in Germany is selected according to specialist categories as language, currency, business objective, geographical location of the company's office, contact and business addresses, telecommunication connections, provider data, advertising banner etc. The search process is further automated by automatically reconciling the qualified URLs and web pages with the data of businesses that are already registered for tax. In addition, the search tool may be used for specific purposes as specified searches for problematic business areas such as 'promoters' of tax evasion and tax havens.

31 The Netherlands started a project at the end of the 1990s, which currently is being further developed in co-operation with the United Kingdom. The project initially concentrated on detecting economic activity, identifying suppliers and extracting relevant information for use in investigations. The use of technology to detect, identify, and monitor e-commerce was studied in detail, in part through the development of prototypes. In view of the knowledge-intensive and technical nature of the project, universities were also asked to take part in this work. As a result of this work, researchers found that good results could be obtained in the fight against illegal or fraudulent e-commerce by using the technology itself to detect Internet-based businesses and to identify the natural or legal persons conducting these activities.

32 The Netherlands search tool may be used as a sampling tool as well as an identification tool. Used as a sampling tool, a quantitative overview can be gained of trends in Internet trading. This is done by automatically collecting and analysing representative samples of websites. Trends can be filtered out by repeating the sampling at fixed intervals.

33 In order to analyse the Internet, a web crawler collects information on websites. The result is a representative sample of Internet sites or a specific collection of sites that satisfy certain selection criteria. For example, one criterion might be that only sites targeted at a particular country may be selected using, for instance, the language of the websites in the selection process. To be representative, the sample must include tens of thousands of sites. All the steps taken to produce this sample of the Internet should be reproducible so that evidence can be furnished and the proper working of the system can be tested.

34 The sample of the Internet is then analysed electronically in order to make a classification of the sites that meet the group characteristics. This classification can be fully automated.

35 To make an analysis, examples of websites are collected that are part of a group of particular interest. The expert system will then identify the distinguishing characteristics and thus determine a site's relevance to the study. It will do so by comparing the content of the sites' textual and HTML features with those of other sites in the sample. The analysis is an iterative process, with the results being assessed by an expert. The expert will review a number of the sites and select positive and negative examples, i.e. sites that in his opinion form part of the target group for the study and sites that (very probably) do not. These examples will be entered into the system again to improve the quality of the results. This will ultimately lead to classified profiles of suspect websites. Positive and negative examples of sites can also be obtained from other sources, such as signals from normal enforcement.

36 For use as an identification tool, a profile of relevant websites is created. This profile is translated into one or more simple search requests on the Internet. Potential identifying data are collected through the use of linguistic technology and extraction techniques.

37 The system localises and extracts characteristic data on the websites concerned such as the postal codes, telephone numbers and VAT numbers. These data are saved in a format that will allow them to be subsequently used for identification purposes. In addition, the system retrieves the website's domain registration data and adds them to the other characteristics. In the identification process the following components are assessed:

- the content of the pages making up the website;
- meta-information on the website;
- the "whois" information of the relevant domain registrant.

38 To optimise performance, the pages that are most likely to contain identification data are accessed first. The process will be ended once enough characteristics have been found to identify the natural or legal person that is responsible for the website. An attempt to identify the supplier electronically will then be made by searching for the same characteristics in relevant data collections. The tax administrations own database is a useful source of information, but many other, external databases are consulted as well. If the supplier cannot be identified electronically, the information from the site is interpreted by an expert and additional investigation will be conducted.

39 Both the German and Netherlands search tools use web crawler technology. The web crawler starts with several good examples of web pages that are interesting from a revenue authority's perspective. The contextual relationship between words on these pages is determined by the use of algorithms which focus the search for similar pages. The object is to scale down the large number of selected web pages that would have been found with a standard web based search engine. The German and Netherlands search tools are capable of getting indications of underreporting but are not able to measure the size of underreporting.

40 In both countries, the use of advanced web search technology is in its early stages and not yet fully integrated into the risk management process of the respective revenue authorities. Both organizations have gained some experience and the first preliminary results indicate that their Internet search tools can be very efficient and effective in the process of detecting situations of potential non-compliance.

Application of the Internet data

41 The web search tools provide revenue authorities with valuable information on activities of economic operators. Depending on a revenue authority's system of risk management and assessment, the information gathered through search tools is processed centrally or provided to local tax offices. In the latter situation, field offices select cases with the highest potential for non-compliance and include a review of Internet activities and revenues generated.

42 The general intelligence captured by industry sector is also useful for documenting the technological environment, the business practices, to enhance training materials, to share best practices, and to develop and describe successful audit strategies.

2 FUNCTIONAL COMPONENTS OF INTERNET SEARCH TOOLS

43 Revenue authorities are increasingly recognizing the special compliance risks that are related to cross border trade over the Internet, risks which are increasing as the trade over the Internet shows a rapid and steady growth. In order to deal with these compliance risks, many revenue authorities have taken organizational steps as a result of which special attention is being devoted to the investigation and monitoring of Internet trade.

44 The information gathered from the Internet is partly of a generic nature, partly case specific. Revenue authorities frequently apply basic Internet search technology - web based search machines like Google and Yahoo - to gather information on Internet trade. Often, the search results are investigated manually. Preferably, more advanced technology should be applied which makes use of the latest searching technology whereby the search process is largely automated. In the operation of such Internet search tools, three main functional components can be distinguished: (1) monitoring the Internet trade and analysing the trends, (2) the collection of high-risk websites and (3) the identification of Internet suppliers.

(1) Monitoring Internet trade

45 As mentioned earlier, country experiences indicate that many revenue authorities are giving greater r attention to monitoring the international trade of goods and services. The search for general information is focused mostly on certain economic sectors that operate over the Internet and pose extra compliance risks. For the collection of general information about trends in Internet trading, most organizations turn to specialised data providers, both governmental and commercial.

46 Understanding must be gained of the patterns of trade over the Internet, in particular the various types of Internet activity and the scale of each type. One of the questions to be answered concerns how certain types of potentially high risk e-commerce activities can be identified. It is necessary to recognize the distinguishing features of each type of activity. Furthermore, the development of various types of activity should be monitored, so predictions can be made of which types will evolve and will disappear.

47 The Internet is a complex research subject with millions of websites as a potential objective for research. Accordingly, one must first acquire a representative, comprehensive, and structured understanding of the Internet, before the identification of specific target groups within the total population of websites is possible. Internet search tools could play an important role in acquiring such understanding by the use of their monitoring function.

48 To gain an understanding, the content of the Internet sites needs to be analysed. Given the enormous size and unstructured nature of the Internet and the speed at which it is changing, it is not feasible to take the entire Internet as starting point for

such analyses. Rather, a relatively small subset of Internet sites should be sampled. This subset will, nevertheless, be very large in absolute terms as it must be sufficiently representative to reveal trends and allow analyses to identify the characteristics of certain types of Internet activities.

49 The sample should be analysed to identify Internet activities that are of interest. The analysis will be confined to a subset of websites with common characteristics that make up a target group. Characteristics of a target group can be profiled by industry (e.g. travel, adult entertainment, gambling, pharmacies, illegal trading), by geography (i.e. local, regional, national, international), by subject (e.g. drugs, weapons) or other relevant criteria.

50 In addition to understanding the current content of the Internet, it is necessary to gain insight into the on going developments that take place. A powerful means to gain such an insight is to compare the results of such analyses of the Internet over time. This can be done in two ways. First, profiles can be drawn up over a relatively long period of time using the same sample from the Internet. These profiles will change over time as the contents of the pages on the websites change. It will then be possible to identify trends within existing websites. Second, the various samples taken of the Internet can be compared with each other over the months and years. The profiles that are drawn up from the samples can also be used to make these comparisons. By monitoring profiles over time, tax officials may gain insight into the development of Internet activities, such as the creation of new activities and the evolution of existing web activities.

51 This trend analysis is not only relevant to gaining an understanding of the developments on a macro level but it can also be used to monitor the behaviour of website suppliers. Changes in the supply of goods and services by individual traders on the Internet can also be of interest. Records can be kept of the kind of e-commerce conducted by a trader, whether the trade is conducted through different websites, and the scale of the trade over time, etc.

52 Trend analysis is recognized as an important tool for revenue authorities to monitor developments in Internet trade. Revenue authorities are encouraged to develop operational software to better understand Internet trade, and thereby improve their capacity to react to quickly to associated high risk non-compliance, both through legislative and administrative responses.

(2) Collection of high-risk websites

53 As mentioned in Part B, country experiences indicate that the methods used by some revenue authorities to collect case-specific information are still quite basic and labour-intensive. More advanced methods, used by relatively small number of organizations, utilize more advanced searching technology where the search process is largely automated.

54 Where it is decided to investigate a group of websites on the Internet, all relevant websites should be identified. Within the domain of e-commerce activities the selection can be refined, for example, by selecting only e-commerce that is targeted at a particular country. Investigations can concentrate on types of activity that were previously identified during the monitoring process. Traditional search engines are obviously inadequate for such complex research issues. A problem with such search engines is that the user must personally define the object of the search. The completeness, representativeness and quality of the search with these machines will remain doubtful.

55 Experience and professional judgement are equally good sources for the manual preparation of profiles to serve as input for the search process. The object in each case is to find all sites on part of the Internet that more or less match a risk profile. Here "part of" means the search process can be restricted to a particular point of interest, for example, a country, while "more or less" refers to the fact that some of the websites found will not match the profile in full but will share some if not most of the main characteristics. How closely the sites should match the profile should be set as a parameter in the search process.

56 The sites found should be ranked by the extent to which they share the characteristics in the profile. When enough sites have been found, the search can be stopped. The result of this is that a group of sites is selected that closely match the profile. The results of this search process consist of the sites found on the one hand and the actual number of sites on the other, which can provide valuable information on the scale of the risk in question. In this respect, changes in the quantity of goods and services being offered on a website may be relevant: are there occasional peaks or is the supply stable or steadily increasing?

57 Revenue authorities are encouraged to develop operational software as a valuable means to collect high risk websites in a cost-efficient manner.

(3) Identification of Internet supplier

58 As mentioned in Part B, country experiences indicate that many revenue authorities still investigate the selected website addresses manually.

59 Where it is decided to check, for example, whether a particular website that supplies goods or services over the Internet, is legal and remits taxes and duties in full, the person liable to tax should be identified.[3] In addition, it might be useful to examine the e-commerce activities of each identified supplier to determine what e-commerce activities are performed over time. Illegal or fraudulent Internet-based activities could thus be detected that would otherwise be overlooked on account of the frequent changes in the goods and services supplied and in the sites used or on account of other variations.

60 In most cases it will be possible to identify the supplier from the information available on a website. Ideally, there should be a way to localise and extract characteristic data on the websites concerned such as the postal codes, telephone numbers, and VAT numbers. These data are saved in a format that will allow them to be subsequently used for identification purposes. In addition, the system should be capable of retrieving the website's domain registration data and add them to the other characteristics. In the identification process, the following components can be assessed:

- the content of the pages making up the website;
- meta-information on the website;
- the "whois" information of the relevant domain registrant.

61 To optimise performance, the pages that are most likely to contain identification data should be accessed first. The process will be ended once enough characteristics have been found to identify the natural or legal person that is responsible for the

[3] Reference is made to the paper on Business Identification published in the OECD Tax guidance series in which the issue of identification is thoroughly analysed. See http://www.oecd.org/dataoecd/4/56/14990201.pdf

website. An attempt to identify the supplier electronically will then be made by searching for the same characteristics in relevant data collections. A revenue authority's own database may also be a useful source of information, but external databases can be consulted as well. Where a supplier cannot be identified electronically, the information from the site should be interpreted by an expert and additional investigations conducted.

62 Revenue authorities are encouraged to develop advanced operational software in accordance with the functionalities described as a valuable means to identify Internet suppliers.

Collaboration in development

63 Country experiences indicate that many revenue authorities recognize the need to counter the compliance risks arising in an e-commerce environment. Collaboration in the development and sharing of the most efficient Internet search tools and techniques is of considerable interest and should benefit all participating revenue authorities. In this respect, it is relevant that some countries are developing or have developed advanced Internet search tools. Due to the characteristics of Internet trade, the information gathered on a macro level by one revenue authority may be of interest to other revenue authorities. Through close co-operation, revenue authorities might therefore learn from each other's successes and failures.

64 The potential application of advanced Internet search tools might also be useful to other governmental investigative services, such as the police and the customs authorities.

65 Developing and using Internet search tools with a view to finding relevant and timely information may help to reduce administrative costs and reduce timeframes as the work is done by computers on an on-going basis.

66 Revenue authorities are encouraged to collaborate in the development of efficient Internet search tools and techniques and to exchange experiences in the use of these tools and techniques.

GUIDANCE

- Revenue authorities are encouraged to develop internet search tools in accordance with the functionalities described to: 1) monitor the Internet trade and analyse the trends; 2) collect high-risk websites; and 3) identify Internet suppliers.
- Revenue authorities are encouraged to collaborate in the development and sharing of the most efficient Internet search tools and techniques and to exchange experiences in the use of these tools and techniques.
- Revenue authorities are encouraged to collaborate in the development of efficient Internet search tools and techniques and to exchange experiences in the use of these tools and techniques.