

## Abstract

Disability insurance claim fraud is costly. Insurance companies have instituted Special Investigation Units staffed with experienced analysts and investigators to combat this problem. There are many tools available for utilization by disability insurance fraud investigators, but few as comprehensive or as valuable as the Internet. Even though this resource is freely available and widely acclaimed, there is a dearth of quality training materials on the subject. While investigators and analysts frequently receive training in the use of other investigative techniques and tools, the Internet represents a formidable challenge. Recognizing the challenge investigators face, this project provides a comprehensive training manual and reference guide for investigating suspicious disability claims via the Internet. More than just a list of helpful websites, the Online Investigator's Handbook provides a methodology for conducting investigative research on the Internet, including Web searching tools and techniques. Additionally, common investigative techniques have been adapted for use in an online environment. The value in utilizing a free resource for investigative research is apparent as companies strive to cut costs and improve efficiency. Special Investigation Units, which may be considered a cost center, are under particular financial strain and will benefit greatly by being able to effectively utilize free or low cost investigative tools. In an era where ecommerce is the norm, traditional investigative techniques such as interviewing and surveillance are necessary but no longer sufficient. In order to conduct a comprehensive investigation, fraud investigators need to be well-versed in online research. The Online Investigator's Handbook provides the instruction necessary to give fraud investigators the required level of online investigation proficiency.

THE ONLINE INVESTIGATORS HANDBOOK:  
A COMPREHENSIVE GUIDE TO ONLINE RESEARCH FOR DISABILITY  
INSURANCE FRAUD INVESTIGATORS

by

James D. Ruotolo

A Project Submitted to the Faculty of  
Utica College

December 5, 2003

In Partial Fulfillment of the Requirements for the Degree  
Master of Science in Economic Crime Management

Members of the Advisory Committee:

Gary R. Gordon, Ed.D., Visiting Professor, Utica College, Utica, NY – Senior Thesis  
Advisor

Thomas Bock, M.S., KPMG – Professional Reader

Mary (Mik) Kelly, The Hartford – Professional Reader

© Copyright by James D. Ruotolo, 2003.

[ This page intentionally left blank. ]

## Table of Contents

<a href="#">Preface</a> .....	1
<a href="#">The Disability Insurance Fraud Landscape</a> .....	1
<a href="#">Statement of the Problem</a> .....	2
<a href="#">Definition of Terms</a> .....	3
<a href="#">Project Rationale</a> .....	5
<a href="#">Introduction</a> .....	17
<a href="#">Assumptions</a> .....	17
<a href="#">How to Use this Handbook</a> .....	17
<a href="#">Online Investigator's Handbook Rationale</a> .....	18
<a href="#">Website Components</a> .....	7
<a href="#">About</a> .....	7
<a href="#">Frequently Asked Questions</a> .....	8
<a href="#">Privacy Policy</a> .....	10
<a href="#">Terms and Conditions</a> .....	12
<a href="#">Other</a> .....	16
<a href="#">Internet Investigation</a> .....	20
<a href="#">Introduction</a> .....	20
<a href="#">Data vs. Information</a> .....	20
<a href="#">Information Classification</a> .....	21
<a href="#">Caveats</a> .....	23
<a href="#">Validating Internet Sources</a> .....	25

<a href="#"><u>Domains</u></a> .....	27
<a href="#"><u>Anatomy of a URL</u></a> .....	27
<a href="#"><u>Determining Domain Ownership</u></a> .....	29
<a href="#"><u>Recovering Missing Webpages</u></a> .....	33
<a href="#"><u>Searching the Web</u></a> .....	36
<a href="#"><u>Browsing</u></a> .....	36
<a href="#"><u>Search Tools</u></a> .....	36
<a href="#"><u>Search Methodology</u></a> .....	39
<a href="#"><u>Common Searching Mistakes &amp; Strategies to Avoid</u></a> .....	40
<a href="#"><u>Advanced Search Techniques</u></a> .....	43
<a href="#"><u>Power Searching</u></a> .....	45
<a href="#"><u>The Deep Web</u></a> .....	48
<a href="#"><u>Topography of the Web</u></a> .....	48
<a href="#"><u>Navigating the Deep Web</u></a> .....	48
<a href="#"><u>Searching the Deep Web</u></a> .....	49
<a href="#"><u>News and Newsgroups</u></a> .....	51
<a href="#"><u>Using News in Investigations</u></a> .....	51
<a href="#"><u>Accessing News Resources</u></a> .....	52
<a href="#"><u>Newsgroup Basics</u></a> .....	52
<a href="#"><u>Accessing Newsgroups</u></a> .....	53
<a href="#"><u>Records Research</u></a> .....	55
<a href="#"><u>Public Records Overview</u></a> .....	55
<a href="#"><u>Telephone Numbers</u></a> .....	55

<a href="#"><u>Addresses</u></a> .....	56
<a href="#"><u>Property records</u></a> .....	57
<a href="#"><u>Driver Licenses &amp; Motor Vehicle Registrations</u></a> .....	57
<a href="#"><u>Professional Licenses</u></a> .....	58
<a href="#"><u>Criminal Records</u></a> .....	59
<a href="#"><u>Business Records</u></a> .....	59
<a href="#"><u>Organizations and Associations</u></a> .....	61
<a href="#"><u>Non-Profit Organizations</u></a> .....	61
<a href="#"><u>Sports</u></a> .....	61
<a href="#"><u>Industry Trade Groups</u></a> .....	62
<a href="#"><u>Hobbies</u></a> .....	62
<a href="#"><u>Anonymous Investigation</u></a> .....	64
<a href="#"><u>Electronic Footprints</u></a> .....	64
<a href="#"><u>Cookies</u></a> .....	64
<a href="#"><u>Online Anonymity</u></a> .....	65
<a href="#"><u>Anonymous Email Communications</u></a> .....	66
<a href="#"><u>Establishing an Undercover Online Persona</u></a> .....	67
<a href="#"><u>References</u></a> .....	72

Preface

*The Disability Insurance Fraud Landscape*

Insurance fraud is costly. While there are several factors that make the actual cost of fraud difficult to quantify, a recent study estimates the annual losses attributed to insurance fraud at \$96.2 Billion (Conning & Company, 2000). Disability insurance represents one of the smallest fraud-loss sectors but still accounts large dollar losses at approximately \$1 Billion each year (Conning & Company, 2000).

Due to the nature of the disability insurance process, long term disability insurance is extremely susceptible to the soft fraud of malingering, where a claimant extends the duration of a disability claim by pretending to be incapacitated in order to receive benefits to which the claimant is no longer entitled. Special Investigation Units (SIUs) have been established by insurers to detect, investigate, and prevent fraud. SIUs have proven effective for insurers and some companies report SIU Return-on-Investment ratios as high as 27 to 1 (Conning & Company, 2000).

By identifying fraudulent claims earlier in the claim lifecycle, insurance companies can become more efficient at investigating fraud and realize greater savings allowing them to focus more energy and resources on the claimants who truly deserve benefits. Fraudulent claims cost insurers billions of dollars each year. Those costs are passed on to the consumer by way of higher premiums. Effective anti-fraud programs can help keep claim costs down allowing insurers to stay competitive and helping individuals and employers to obtain disability insurance at a reasonable rate.

*Statement of the Problem*

Increasingly, investigators are turning to the Internet to begin their research or investigation into a suspicious claimant. The Internet affords myriad resources to the investigator with minimal time or effort. Investigators utilize all kinds of online tools including electronic public record services like ChoicePoint or LexisNexis, insurance claim-specific services like ISO ClaimSearch, and others. These paid online resources offer excellent information and are accessible from the desktop. Insurers recognize that these technological tools are essential to effectively combat fraud. A 2000 survey of insurance companies found that companies are investing in technological solutions both to identify suspicious claims and build a case to prosecute suspected offenders (Conning & Company, 2000). Additionally, Nearly 90% of survey respondents felt that the use of technology could improve their fraud detection efforts (Conning & Company, 2000).

In addition to the common subscription resources, there is another resource that investigators have discovered hiding in plain sight. The World Wide Web itself can provide a wealth of information about a claimant's activities or whereabouts. Using Web resources effectively is highly desirable because they are free, fast, and effective. However, while most SIUs have been provided with a computer and Internet access, very few have received sufficient formal training on how to use Web resources to conduct investigations.

While there are some training materials available on this topic, there is a dearth of comprehensive quality training on this subject. This void has been haphazardly filled with some self-study courses, seminars, and other inadequate training tools. However, there is very little published material available on how to effectively use the Internet as a resource for the investigation of insurance fraud claims. Disability insurance claim investigators in

particular can benefit from more formal training in this area, as disability fraud investigations often center around functionality and activity – the types of information that are readily available on the Internet.

This handbook provides a comprehensive training guide for disability insurance fraud investigators and analysts in using no cost (or low cost) Internet resources to supplement traditional investigative techniques.

### *Definition of Terms*

Internet. A decentralized global network connecting millions of computers (Webopedia, 2003).

World Wide Web. A system of Internet servers that support the HyperText Markup Language (HTML) (Webopedia, 2003). Also referred to as “WWW” or “Web.”

Deep web. World wide web content stored in databases accessible on the Web but not available via search engines. Also called the “Invisible web” since content is invisible to search engines. (Cohen, 2003).

Special Investigation Unit (SIU). Entity responsible for investigating suspicious claims on behalf of an insurer. The SIU function may be provided by an internal group or outsourced to a third-party. Some states specifically require that this entity be separated from both claim and underwriting departments.

Insurance claim. Demand for payment in accordance with an insurance policy (WordNet, 2003). For the purposes of this project, “claim” will refer to an insurance claim in accordance with a disability insurance policy unless otherwise noted.

Investigation. A thorough and systematic inquiry conducted by an insurers Special Investigation Unit, law enforcement authority, or other investigative body into questionable activities surrounding a disability insurance claim. Note that “claim investigation” is an insurance industry term commonly used to define the regular course of business when processing a claim. In this project, Investigation refers only to special inquiries made due to suspicious activity and not regular claim processing procedures.

Definition of Disability. The policy provision establishing the conditions for eligibility of benefits. The two major categories of long term disability are Own Occupation and Any Occupation.

Own Occupation. In order to meet the Definition of Disability under this contract provision, a claimant due to sickness or injury must be unable to perform the substantial and material duties of his or her regular occupation (Delizia, 2003).

Any Occupation. In order to meet the Definition of Disability under this contract provision, a claimant due to sickness or injury must be unable to perform the substantial and material duties of any occupation for which he or she is suited by education, training and experience (Delizia, 2003).

Proof of loss. A formal statement made by a claimant to an insurer regarding a loss. In the case of a disability claim, a claimant must provide proof that he or she is disabled under the provisions of the policy. Generally, proof of loss must be provided when a claim is filed and depending on the contract language, an insurer may ask for additional documentation to determine if a loss has occurred.

Malingering. Extending the duration of a disability claim by pretending to be incapacitated in order to receive benefits to which the claimant is not otherwise entitled. Also known as “soft fraud.”

Publicly available. Information available to the general public from non-governmental sources such as telephone directories, classified ads, newspaper articles, periodicals, or other forms of publicly distributed information.

Public record. Information about or related to an individual obtained originally from the records of a federal, state, or local governmental entity open for public inspection.

Non-public. Information about an individual of a private nature neither available to the general public nor obtainable from a public record. Access to Non-Public information generally requires permission either granted by statute or by authorization from the individual.

### *Project Rationale*

This project aims to provide a formal training guide for investigators. Almost all SIUs provide access to the Internet for their investigative staff and few would argue that the Internet is a powerful resource. However, very few investigators have an extensive background in library science, Boolean logic, or software development. While some companies and agencies do offer computer training, few if any of these are geared toward Internet research.

Investigators have tended to use the highly reliable paid search tools almost exclusively, both because of their reliability and because of a sales force that is willing to train new users. The Internet does not have a sales force to provide training and these paid

services are under increasing pressure from privacy groups to limit the information available to users. Suits against public record aggregators represent an emerging trend in litigation (Christensen, 2003). Finding alternative investigative resources is wise. The public Internet is generally immune from such litigation in that it is inherently provided for public consumption. Some SIUs are viewed as cost centers and, with the persistent goal of reducing expenses, the World Wide Web presents an attractive option as a research and investigation tool as the information available is largely free or available at minimal cost.

## Website Components

### *About*

The Project. The Online Investigator's Handbook was created as a part of an educational research project, in partial fulfillment of the requirements for a Master of Science degree in Economic Crime Management from Utica College. The author designed the handbook to be both a comprehensive self-paced tutorial and a reference guide for online investigative research. Investigators working on cases that might benefit from the use of online investigative tools can use the handbook to improve their online research skills and learn new ways to use the Internet for investigative purposes.

The Handbook is more than a list of helpful links. This project introduces a new investigative methodology for using online tools effectively to investigate fraud, conduct background checks, enhance due diligence, and research complex topics. The Online Investigator's Handbook combines this methodology with practical advice, real-world examples, and cutting-edge resources to give investigators the tools they need to make the most out of the wealth of information available in today's online world.

The first edition of the Online Investigator's Handbook, released in December 2003, focused on the investigation of disability insurance claim fraud. However, the methodologies discussed and tools provided throughout the handbook are applicable to almost all types of investigations.

The Program. Utica College has been the North American leader in economic crime study since 1988, when the Economic Crime Investigation undergraduate major was first offered. Working with the Board of Directors of the Economic Crime Institute, Dr. Gary R.

Gordon founded the Economic Crime Management (ECM) master's degree program in 1999. The ECM program, offered in a distance learning format with three one-week long residencies each year for two years, is offered at Utica College in Utica, New York.

The ECM curriculum focuses on fraud and risk management strategies, current economic crime challenges including money laundering and identity fraud, and applying innovative technological and analytical solutions. It has been designed to meet the growing security demands placed on experienced personnel in law enforcement, private corporations, government, and the military. For more information about these programs, please visit the Utica College and Economic Crime Institute websites.

The Author. James D. Ruotolo currently works at a major life and disability insurance company as a Senior Business Analyst in the Special Investigation Unit. Mr. Ruotolo's responsibilities include project management, new technology development, and anti-fraud training. His public speaking engagements include training sessions on Online Research and Investigation Tactics for the National Healthcare Anti-Fraud Association (NHCAA), International Association of Law Enforcement Intelligence Analysts (IALEIA), and the International Association of Special Investigation Units (IASIU). Mr. Ruotolo holds a Bachelor's degree in Economic Crime Investigation and is a Masters degree candidate in the Economic Crime Management program at Utica College.

### *Frequently Asked Questions*

What is the Online Investigator's Handbook? The Online Investigator's Handbook is a training tutorial and reference guide for investigators. It presents a new methodology for

conducting research on the Internet when used as part of a fraud investigation. It began as a graduate research project. For more information, see the About section.

What do you mean by "online" investigations? "Online" investigation tactics can be used for all types of crimes or research - not just computer fraud. The handbook started as a tool for the investigation of fraudulent disability insurance claims but the procedures apply to all types of investigations. Basically, "online investigations" means using a computer to find out information on a subject.

Is this the same thing as computer forensics? No. Computer forensics generally involves the analysis of data on a specific computers. For example, computer forensics technicians might examine a hard disk drive on a computer seized during a search of a suspect's home in order to identify evidence of a crime. Although it is a powerful investigative tool, the Online Investigator's Handbook does not cover this topic.

How do I report a broken link? The dynamic nature of the Internet is one of its greatest assets but it also causes problems for websites with multiple links to other pages. If you come across a broken link in the handbook, please report it by visiting the Contact page.

I think I can use online information in my investigation but I need more help. Who should I call? First, read the handbook again. Chances are, the answer to your question is in there. If you need more assistance, contact the Online Investigator's Handbook administrator with a specific question. Please note that due to the volume of requests, the administrator may not be able to respond to all requests but every attempt will be made to provide assistance when possible.

Don't the paid public record services give you the same information, all in one place for a small fee? Some of the information discussed in the handbook is available from public

record aggregators. In fact, sometimes the author recommends using these types of paid services for particular types of information because of their high level of accuracy. However, much of the information discussed in the handbook is not available from such paid services. The handbook focuses heavily on how to use the free resources available to conduct investigations.

I know of a resource that was not mentioned in the handbook. Do you take recommendations? The Internet is in a constant state of change and new resources pop up all the time. If you have a website or method that you would like to recommend, please visit the Contact page. Remember to include your name and email address!

Where can I find more information? Hopefully, the Handbook has increased your interest in this topic. Check out the References section for additional sources of information. Also, you may wish to review the Tools & Resources page for more ideas.

### *Privacy Policy*

Introduction. The Online Investigator's Handbook is strongly committed to protecting our users' privacy. As part of that commitment, we have outlined in detail what information will be collected, what measures are in place to protect that information, how that information will be used, with whom that information may be shared and on what basis that sharing will take place. And we assure you, the user, that we will at all times abide by the terms of this Privacy Policy and honor the trust that you have invested in us. If you have any questions or concerns regarding this Privacy Policy, please contact us; we will do our best to address your query in a timely and professional manner.

Legal Compliance. OIH may share information in order to investigate, prevent, or take action regarding illegal activities, suspected fraud, situations involving potential threats to the physical safety of any person, violations of OIH's terms of service, or as otherwise required by law.

Links. OIH provides links to other Websites as a resource to users. OIH is not responsible for the privacy practices of other sites, and no representations are made as to personally identifiable information gathered by those Websites themselves. We encourage our users to be aware when they leave our site and to read the privacy statements of each and every web site that collects personally identifiable information. The existence of a link to another Website does not imply that such Website or its owners and operators are affiliated in any way with OIH.

Information Collection. OIH may utilize traffic statistics as a tool to optimize our site's usefulness to our users. For this reason, as well as for security purposes, OIH may collect I.P. Addresses, browser type, domain names, access times, referring URL, and page view information of users who browse all areas of our Website. Individual information will not be shared with third parties except as necessary for legal compliance. Aggregate site statistics may be shared for evaluation purposes. OIH does not collect individually identifiable information, except that which is voluntarily provided by users via the Contact page or email communication.

Notification of Changes. OIH will occasionally update this Privacy Policy. It is our commitment that our users will always be aware of what information we collect, how we use it, and under what circumstances we disclose it. OIH encourages you to periodically review this Policy to be informed of how we are protecting your information.

*Terms and Conditions*

Terms of Service. By utilizing the Online Investigator's Handbook Website, you agree to be bound by the following terms and conditions ("Terms of Service"). If you do not agree with any term or condition set forth below, you are not authorized to use the services for any purpose. The Online Investigator's Handbook (OIH), its authorized agents, and author James D. Ruotolo, are herein described as "We."

Personal Use Only. OIH is made available for your personal, non-commercial use only. You cannot distribute or transfer copies to others in exchange for money or other consideration. You may not use OIH to sell a product or service, or to increase traffic to your Web site for commercial reasons, such as advertising sales.

You may not modify, reformat, copy, display, distribute, transmit, publish, license, create derivative works from, transfer or sell any information, products or services obtained from OIH, or mirror the website home page or handbook pages on your Website.

If you want to make commercial use of OIH, you must enter into an agreement with OIH to do so in advance. Please contact us for more information.

Changes In Terms of Service. We may modify or terminate our services from time to time, for any reason, and without notice, including the right to terminate with or without notice, without liability to you, any other user or any third party. We reserve the right to modify these Terms of Service from time to time without notice. Please review these Terms of Service from time to time so that you will be kept abreast of any changes.

Copyright and Trademark. All contents of OIH are Copyright © 2003-2004 James D. Ruotolo. All rights reserved. The Online Investigator's Handbook logo, trademarks and

other marks that appear throughout OIH belong to James D. Ruotolo, and are protected by U.S. and international trademark laws. You are prohibited from using any of the marks appearing throughout OIH without express written consent from James D. Ruotolo, as permitted by applicable laws.

Companies, products, websites, and other entities mentioned in the Online Investigator's Handbook may be trademarks or registered trademarks of their respective trademark owners. The listing of a company, product, website or other entity in the Online Investigator's Handbook does not indicate an affiliation with the Online Investigator's Handbook. Unless stated explicitly, no endorsement of this information, service or product by any company or person is made or implied.

Disclaimer of Warranties. We disclaim any and all responsibility or liability for the accuracy, content, completeness, legality, reliability, or operability or availability of information or material found within. We disclaim any responsibility for the deletion, failure to store, misdirected delivery, or untimely delivery of any information or material. We disclaim any responsibility for any harm resulting from downloading or accessing any information or material on the Internet through OIH.

OIH is provided "as is," with no warranties whatsoever. We expressly disclaim to the fullest extent permitted by law all express, implied, and statutory warranties, including, without limitation, the warranties of merchantability, durability and fitness for a particular purpose, quality, satisfactory quality, title and non-infringement of proprietary rights. We disclaim any warranties regarding the security, reliability, timeliness, and performance of OIH. We disclaim any warranties for any information or advice obtained through OIH. We disclaim any warranties for services or goods received through or advertised on OIH or

received through any links provided by OIH, as well as for any information or advice received through any links provided in OIH.

OIH is intended to be a reference for investigators. It offers recommendations for sources of information and methods of searching. These recommendations are not considered to be best practices and We disclaim responsibility for any damages incurred as a result of the use of these recommendations. The information contained in this website does not constitute legal advice.

All Uniform Resource Locator (URL) hyperlinks included in this document have been verified and were active at the time of production. We cannot be held responsible for changes in hyperlinks, broken hyperlinks, or websites that are no longer active. We are not responsible for the content of any source recommended in the OIH. We make no claims, promises or guarantees about the accuracy, completeness, or adequacy of the information contained in or linked to this web site or any web site, document, reference, presentation, or other material referenced in this web site.

You understand and agree that you download or otherwise obtain material or data through the use of the OIH at your own discretion and risk and that you will be solely responsible for any damages to your computer system or loss of data that results in the download of such material or data.

Some states or other jurisdictions do not allow the exclusion of implied warranties, so the above exclusions may not apply to you. You may also have other rights that vary from state to state and jurisdiction to jurisdiction.

Limitation of Liability. Under no circumstances shall We be liable to any user on account of that user's use or misuse of or reliance on OIH. Arising from any claim relating to

this agreement or the subject matter hereof such limitation of liability shall apply to prevent recovery of direct, indirect, incidental, consequential, special, exemplary, and punitive damages whether such claim is based on warranty, contract, tort (including negligence), or otherwise, even if We have been advised of the possibility of such damages. Such limitation of liability shall apply whether the damages arise from use or misuse of and reliance on OIH, from inability to use OIH, or from the interruption, suspension, or termination of OIH (including such damages incurred by third parties). This limitation shall also apply with respect to damages incurred by reason of other services or goods received through or advertised on OIH or received through any links provided in OIH, as well as by reason of any information or advice received through or advertised on OIH or received through any links provided in OIH. This limitation shall also apply, without limitation, to the costs of procurement of substitute goods or services, lost profits, or lost data. Such limitation shall further apply with respect to the performance or non-performance of OIH or any information or merchandise that appears on, or is linked or related in any way to, OIH. Such limitation shall apply notwithstanding any failure of essential purpose of any limited remedy and to the fullest extent permitted by law.

Some states or other jurisdictions do not allow the exclusion or limitation of liability for incidental or consequential damages, so the above limitations and exclusions may not apply to you.

Without limiting the foregoing, under no circumstances shall We be held liable for any delay or failure in performance resulting directly or indirectly from acts of nature, forces, or causes beyond its reasonable control, including, without limitation, Internet failures, computer equipment failures, telecommunication equipment failures, other equipment

failures, electrical power failures, strikes, labor disputes, riots, insurrections, civil disturbances, shortages of labor or materials, fires, floods, storms, explosions, acts of God, war, governmental actions, orders of domestic or foreign courts or tribunals, non-performance of third parties, or loss of or fluctuations in heat, light, or air conditioning.

Choice of Law. These Terms of Service will be governed by and construed in accordance with the laws of the State of Connecticut, without giving effect to its conflict of laws provisions or your actual state or country of residence. If for any reason a court of competent jurisdiction finds any provision or portion of the Terms of Service to be unenforceable, the remainder of the Terms of Service will continue in full force and effect.

Entire Agreement. These Terms of Service constitute the entire agreement between the parties with respect to the subject matter hereof and supersedes and replaces all prior or contemporaneous understandings or agreements, written or oral, regarding such subject matter. Any waiver of any provision of the Terms of Service will be effective only if in writing and signed by an authorized agent of OIH.

#### *Other*

In addition to the sections described above, the Online Investigator's Handbook website will also contain a homepage with welcoming information, a search function, a contact form for feedback, and a list of credits.

The handbook chapters include: Introduction, Internet Investigation, Domains, Searching the Web, The Deep Web, News and Newgroups, Records Research, Organizations and Associations, and Anonymous Investigations.

## Introduction

### *Assumptions*

The author has made several assumptions regarding the use of the Online Investigator's Handbook website. Users must have access to the Internet via modern browser software (Internet Explorer 5.x, Netscape 6.x, or better) and have a basic understanding of the following topics: Insurance investigation, Internet surfing, email, and general computer usage. The author also assumes that the user has conducted Internet research using search engines or other common Web searching tools. Each search tool on the Internet uses its own syntax and users are encouraged to read the Help sections of each tool before using it. All examples in this handbook are given utilizing the syntax for the Google search engine unless otherwise noted.

Companies, products, websites, and other entities mentioned in the Online Investigator's Handbook may be trademarks or registered trademarks of their respective trademark owners. The listing of a company, product, website or other entity in the Online Investigator's Handbook does not indicate an affiliation with the Online Investigator's Handbook. Unless stated explicitly, no endorsement of this information, service or product by any company or person is made or implied.

### *How to Use this Handbook*

This handbook is designed to be a handy resource for investigators when conducting online research in relation to disability insurance claim fraud investigations. It is not a definitive guide on public records, research methods, or investigative strategy. This guide is

designed to give investigators the background and tools they need to effectively gather online information during a fraud investigation.

The handbook is divided into chapters and each chapter is divided into sub-categories. The handbook may be read like a normal book, chapter-by-chapter, for a complete tutorial on the subject matter. Alternatively, the user may wish to jump to a particular section of the manual and use it as a reference guide when conducting research. Please note that words appearing in fixed width font in this handbook are examples or words that should be typed during a search query.

#### *Online Investigator's Handbook Rationale*

Increasingly, investigators are turning to the Internet to begin their research or investigation into a suspicious claimant. The Internet affords myriad resources to the investigator with minimal time or effort. Investigators utilize all kinds of online tools including electronic public record services like ChoicePoint or LexisNexis, insurance claim-specific services like ISO ClaimSearch, and others. These paid online resources offer excellent information and are accessible from the desktop. Insurers recognize that these technological tools are essential to effectively combat fraud. A 2000 survey of insurance companies found that companies are investing in technological solutions both to identify suspicious claims and build a case to prosecute suspected offenders (Conning & Company, 2000). Additionally, Nearly 90% of survey respondents felt that the use of technology could improve their fraud detection efforts (Conning & Company, 2000).

In addition to the common subscription resources, there is another resource that investigators have discovered hiding in plain sight. The World Wide Web itself can provide a

wealth of information about a claimants activities or whereabouts. Using Web resources effectively is highly desirable because they are free, fast, and effective. However, while most SIUs have been provided with a computer and Internet access, very few have received sufficient formal training on how to use Web resources to conduct investigations.

While there are some training materials available on this topic, there is a dearth of comprehensive quality training on this subject. This void has been haphazardly filled with some self-study courses, seminars, and other inadequate training tools. However, there is very little published material available on how to effectively use the Internet as a resource for the investigation of insurance fraud claims. Disability insurance claims in particular can benefit from more formal training in this area, as disability fraud investigations often center around functionality and activity – the types of information that are readily available on the Internet.

Remember that this handbook is not intended to be a comprehensive list of all the resources available to investigators. The goal of this project is to provide a methodology along with a few representative examples and tools for users enhance their ability to gather information effectively.

## Internet Investigation

### *Introduction*

The Internet is the largest data source on the planet and it can be an superb research tool when used correctly. The problem most people have when they begin to conduct research on the Internet is that they don't know where to start. The key to efficient research in any medium is information management—parsing out the useful data while avoiding superfluous information. This process becomes more and more difficult as the information source gets larger. Thus, the Internet can seem like an insurmountable challenge to a novice researcher.

This handbook is designed to be a resource for investigators when conducting online research in relation to disability insurance claim fraud investigations. It is not a definitive guide on public records, research methods, or investigative strategy. This guide is designed to give investigators the background and tools they need to effectively gather online information during a fraud investigation.

### *Data vs. Information*

Many people use the terms “data” and “information” as synonyms. However, there are differences and these differences are fundamental to conducting Internet research. Data consist of facts. Investigators are accustomed to gathering and reviewing all the available facts. Information results from the analysis and interpretation of data (Dictionary.com, N.D.). Utilizing this definition, it is important to understand that data alone have no meaning. Data are only valuable when they are interpreted, combined with context, and thus turned into

information. As you read through this handbook, remember the following formula: Data + Context + Analysis = Information.

Internet researchers sometimes focus on Data without spending sufficient effort identifying context or conducting analysis. Internet sources can help with each of these items and it is important to remember all of them are required in order to obtain useful and actionable information. Information can be divided into the three major categories (Individual Reference Services Group, 1997) described in the section below.

### *Information Classification*

Publicly available information is defined as information about an individual that is available to the general public from non-governmental sources such as telephone directories, classified ads, newspaper reports, publications, or other forms of information (Individual Reference Services Group, 1997). An Investigator does not need any special permission or authorization to obtain this type of information. Additionally, a large amount of publicly available information is accessible for free on the internet. Many traditional hard copy sources of non-public information such as newspapers and phonebooks now have online counterparts. The advantage of the online version is that a user is able to search the contents of many documents at once as opposed to scouring over pages of articles or numbers by hand. Nearly all major newspapers and many local newspapers have provided users with the ability to search their article archives for free. Searching online for a claimant's name or address in a local paper can provide excellent results. For more information, see the section on News Information.

Public record information is defined as information about or related to an individual which has been obtained originally from the records of a federal, state, or local governmental entity that are open for public inspection (Individual Reference Services Group, 1997). Contingent upon state statutes, many of the records collected by government agencies are available for public viewing. Property deeds, motor vehicle records, liens, judgments, business credit reports, and professional licenses are some examples of public records. Most of these records can be obtained without special authorization or notification to the claimant. Public record information is available from a variety of sources, the most reliable of which is the record generator - the government agency that procured the record. Other sources include data aggregators who collect public records from a variety of primary sources and combine the information in easily accessible databases. Services like ChoicePoint and Lexis-Nexis provide this type of information. These services are less reliable than the original source because they need to collect and update their databases and updates can range from hourly to annually. However, the advantage provided by these types of services is that a variety of information is available in a single place, usually within seconds whereas attempting to get data from a government office can take weeks or months. These services usually require a subscription and all of them are fee-based. There are, however, some public records that have been made available online via the Internet free of charge. In most cases, a government office has decided to provide internet accessibility to a database under its control. Investigators can take advantage of these free services but caution should be used. Free services are notorious for being out of date or otherwise disabled. Information obtained via a free internet service should always be verified by another reputable independent source.

For more information about public records, see the section on Records Research and review the section below on Validating Internet Sources.

Non-public information is defined as information about an individual that is of a private nature and neither available to the general public nor obtainable from a public record (Individual Reference Services Group, 1997). Examples of non-public information are medical records, tax records, employment history, and personal credit reports. In order to legally obtain this information, an investigator must have authorization from the claimant and, for credit reports, a legitimate permissible purpose according to Fair Credit Reporting Act (2002). Non-public information is generally available through a limited number of sources. Examples of these are: the individual claimant themselves, the agency/organization/company who collects the information such as the Internal Revenue Service, doctor, credit agency, or a designated third party such as an attorney.

### *Caveats*

The primary use for the types of records described above is independent verification of claimant provided data. Obviously, these sources can be used when the claimant is hesitant to provide this type of information on his or her own. In this case, multiple independent sources should be consulted where possible. The validity, creditability, and integrity of information is increased dramatically when multiple independent corroborating sources are used.

One of the main reasons for the explosive and continual growth of the Internet is the lack of regulation and control over it. Anyone in the world can post information to the Internet or setup a website (Cohen, 2003, Conducting). This is what makes the Internet the

wonderful resource that it is. Unfortunately, it is also a cause for concern when using the Internet as a research tool. It is often difficult to verify information obtained from the Internet. It is very simple for one person with an outdated home computer to make a website that looks as professional as that of a multi-billion dollar corporation. If someone were to purposely misrepresent themselves, upon cursory examination it would be difficult to determine the subterfuge. It is extremely important to verify any information obtained from an Internet source. Investigators should be prepared to deal with two key issues when conducting research using free Internet sources:

**Information Overload.** There is a huge volume of information available on the Internet and even seemingly obscure topics can have large numbers of relevant websites or documents. Investigators need to be able to evaluate this information expeditiously.

**Validity.** Misinformation abounds on the Internet. Urban legends, fictitious accounts, ghostwriters, and various types of misleading information can confuse the researcher. Since anyone can publish information on the Web, it is important to validate everything. When validating Internet sources, the best method is to use a non-Internet source.

**Dated Material.** The Internet has the ability to deliver information at lightning speed. In the business world today, information flow is quick and constant. Computer users have come to rely on speedy information delivery. Users have grown accustomed to this rapid delivery of information and now there is a notion that all information on the Internet is current, up-to-date, and cutting edge. This is a dangerous and incorrect assumption. The Web is brimming with outdated articles and sites (Jesdanun, 2003). Be careful when reviewing Internet-provided information and make sure to verify the dates! Simply because a source

contains a date does not mean that it is correct. Attempt to verify this information by using the techniques described below.

### *Validating Internet Sources*

There are several strategies that Investigators can employ when evaluating information found on the Internet.

Analyze the Domain. Domains themselves may provide validation information about the source. Top-level domain (TLD) sites are generally more reliable than sites that are published in user communities. For example [www.realbusiness.com](http://www.realbusiness.com) is likely to be more reputable than [www.geocities.com/~johnsmith/useless/index.html](http://www.geocities.com/~johnsmith/useless/index.html) (Barker, 2003, Evaluating). Also, take note of the domain type. Common types are .com, .org, and .edu. For more information, see Anatomy of a URL.

Determine the Author. Not all information posted online contains a byline or other indication of who the author is. It is important for an investigator to know where the information is coming and whether or not that source is reliable. Look for attributions in headings, titles, or at the end of a document. For more tips, see Determining Domain Ownership. Also, when reviewing websites of companies or organizations, look for brick-and-mortar contact information. Website authors who provide brick-and-mortar contact information like physical address and phone number may be more reliable. Information provided by a verifiable authority, such as the Secretary of State, is more reliable than an aggregator or reseller of information.

Identify Bias. Based on the analysis of the domain, determine any bias that the author may have (Kirk, 2002). .com domains generally belong to for-profit corporations. Does the

information contain biased marketing hype? .edu domains are reserved for educational institutions. Has the information been posted by a reputable professor or a undergraduate student? Researchers should ask themselves, why did this person publish this material online? Identifying the motivation for publishing material can help identify possible bias. Biased information is not necessarily bad, but the context must be correctly understood and incorporated into any decision made based on that information.

Scan page perimeter. When seeking identifying information on a website, look to the perimeter of the pages within that site (Barker, 2003, Evaluating). A standard website format will include some identifying banner across the top of the page, links to other pages on the left or right-hand sides, and miscellaneous details at the bottom of the page. Often, links to privacy policies, service listings, resumes, or other useful material will be hidden at the bottom of a page. Don't forget to scroll down!

Additional Methods. There are several other ways to validate internet resources. Contacting the source directly by phone, email, or snail mail may provide the authentication required. Also, use the advanced features of some search engines to determine the popularity of a website. For example searching for link:www.microsoft.com with Google will show you all the sites that link to Microsoft.com. If reputable sites have chosen to link to the target site, this may be a sign of credibility (Barker, 2003, Evaluating). Also, searching for an author or company name in a major search engine may yield information that validates the credibility of the source. For more information about how to do this, see Searching the Web.

## Domains

*Anatomy of a URL*

Understanding the anatomy of a Uniform Resource Locator (URL) is essential when conducting online research. Frequent users have become accustomed to seeing these URLs in their browser's address bar but not everyone knows what they mean. The following URL will be used as an example:



*Figure 1.* A diagram of a sample Uniform Resource Locator (URL) address and its individual components. Copyright 2003 by James Ruotolo.

This single address actually contains five distinct segments and each of them provides some very basic information:

Protocol. In the example described above, the protocol is `http` which stands for Hyper Text Transfer Protocol (Bell, 2003). There are other types of protocols as well. Common ones include `ftp`, and `telnet`. The protocol in a URL is always followed by a colon and two forward-slashes (`http://` or `telnet://` for example). HTTP is the protocol that is used for the overwhelming majority of web content that most users are familiar with.

Investigators may also encounter the `https://` protocol. The addition of an "S" indicates Secure Sockets Layer (SSL) encryption is being used. This protocol is frequently used for secure ecommerce transactions and is accompanied by a key or lock symbol in the lower

right-hand corner of the browser window. For more information on SSL, visit Introduction to SSL, provided by Netscape.

Host server. The next segment identifies the server where the files are hosted (Bell, 2003). Computers recognize host servers via their IP address, not the text based web addresses that most users are familiar with. However, since IP addresses are long strings of numbers and are difficult for humans to remember, a system called the Domain Name Service (DNS) was implemented to automatically translate the textual notations that we use into the IP Address numbers that the computer needs (Webopedia, 2003, DNS).

In our example, the host server is `www.sample.com/`. Note the `.com` extension. This signifies that the website is a Top Level Domain (TLD) of a certain type. The Internet Corporation for Assigned Names and Numbers (ICANN) is responsible for controlling these TLDs and has standard requirements for certain TLD extensions. Identifying the type of TLD is important when conducting investigations. It helps to identify the author and uncover any potential bias that the person has for posting certain information. Companies might have a profit motive while educational websites might post material simply for the research value. The TLD can help an investigator determine why information has been made available on the Internet. Listed here are some common TLDs (Webopedia, 2003, TLD): `.com` = Company, `.mil` = Military, `.org` = Organization, `.gov` = Government, and `.edu` = Educational. For a complete listing of TLD extensions, see [AllWhois.com](http://AllWhois.com).

Directory path. There may be folders located on the host server. If the file that is being accessed resides in a folder, there will be several more forward slashes – each indicating a sublevel in a folder hierarchy. In our example, there are two folders, referenced by `/conferences/2004/`. The file we are accessing resides within the 2004 folder

which is within the conferences folder. Note that punctuation and capitalization count when entering URLs in the address bar. Different directory structures and operating systems running on the web server machine handle capitalization, spaces, and other punctuation differently. Therefore, if you are entering a URL directly into the address bar, as opposed to clicking a link, it is important to type it exactly as it is written. Remember to look for the word “users” or a tilde (~) in the URL. This frequently denotes a website or webpage hosted by an ISP and means that the author of the page most likely does not own the whole TLD (Barker, 2003, Glossary).

Filename. There are many different types of files that can be accessed over the Internet. The most common file types are html files, most often given with the extension .htm or .html. In our example, the URL points to the file `sessions.html`. This file contains the content that is shown in the browser window. Other types of files that might be displayed in the browser include text files (e.g. `sample.txt`), Adobe Portable Document Format files (e.g. `sample.pdf`), and so on.

Bookmark or Anchor. In this example, the URL contains some additional characters not always seen in Internet addresses. The #A in the URL is a bookmark, also known as an anchor (Bell, 2003). The author of the page has created bookmarks within the page to tell the browser to scroll to a certain point on the webpage. The title of the bookmark, "A" in our example, is set by the webpage author.

### *Determining Domain Ownership*

Whois. It is not uncommon for a claimant who has access to the Internet to have his or her own website. Increasingly, Internet Service Providers (ISPs) offer free website

hosting as part of internet service plans. Many times individuals will include an “About Me” section which generally provides personal details and photos. Investigators can find useful information about functionality and daily activities in these areas. One claimant even had color photos of his surgery posted on his own website! Claimants operating businesses may have a website for the business as well. Identifying unreported income or functionality that is inconsistent with the stated limitations associated with a particular diagnosis is key information during a disability claim fraud investigation. If a website suspected to belong to the claimant is encountered, it is important to determine who the true owner of the website is.

Now that we know what the components of a URL are, how do we find out who owns a website? Sometimes, the domain owner seems obvious. Everyone knows that the microsoft.com domain belongs to Microsoft Corporation. However, making assumptions here is dangerous. For example, many unsuspecting users have visited the whitehouse.com domain looking for information about the President but found a website featuring adult content instead.

Fortunately, there is a way to verify the owner of a website. Whois (pronounced "who is") utilities provide information about the owners of Top Level Domains (TLDs). There are many whois services and users are encouraged to try several when attempting to verify registration information. As an example, searching for the domain smith.com with Network Solutions Whois, returns a wealth of information.

In this example, Smith International, Inc. is the registered domain owner and contact information is also provided. An administrative contact is listed. Note that the administrative contact person is from the same company. This may not always be the case. Sometimes, a company or organization may outsource the administration of their website to

another agency. However, this information is usually representative of the true owner and person responsible for the website content. This is key information for the investigator. Also take note that, in this case, an email address is provided.

A technical contact is also listed. This person or group is generally responsible for the actual hosting of the website and may often be an ISP. Frequently, this person or agency may not be responsible for the content of the site in any way but rather is only responsible for hosting the site on a web server.

Dates of creation and registration expiration along with domain server names are noted near the bottom of the record.

Other popular whois services include AllWhois, and FasterWhois. Most of these search the common TLDs like .com, .net, and .edu. Some domains are restricted and require their own whois search. Visit the ICANN website for more information on how to search restricted domains.

Hidden HTML. Also, it may be useful to check out the html code for a webpage. Sometimes web developers insert comments which are generally hidden from view unless the actual code is displayed. These comments are generally used for development purposes – notes to other developers, references to a particular programming method, and so on. However, they may contain references to the author. To view the html code for the webpage currently being viewed in a browser, select Source from the View menu.

To identify comments in html code, look for the Less Than and Exclamation Point characters, usually followed by two dashes. For example `<!-- Hello -->` is an html comment. The `<!` symbol denotes the beginning of the comment. The text in between the

Less Than and Greater Than brackets is hidden from webpage viewers but is visible in the actual html code.

Html comments do not always contain useful intelligence but they may contain helpful identifying information and are worth a quick look. When viewing the code, use the Find feature on the Edit menu to search for <! to find comments buried in the code.

File Properties. Another simple method of getting some historical information about a webpage is simply checking the properties of the html document that is being viewed. If a user is looking at a webpage in a browser, the page properties can be viewed by selecting Properties from the File menu in Internet Explorer or selecting Page Info from the View menu in Netscape Navigator.

The information contained in the properties box includes the date the file was created and last modified. These dates may be useful in determining when a person was actively involved in editing or changing a website and also when to go back and view a historical copy of a website. However, be careful when using this information as it may not be accurate (Barker, 2003, Evaluating).

Other Techniques. There are some less official methods for checking the identity of a website or webpage owner. Certain sections of a website are likely to list information about the author. Obviously, investigators should look for “About Me” or “Resume” pages. Also be on the lookout for identification information on the bottom of pages. Website created / designed / maintained / hosted by, etc. information could be useful in identifying the person who maintains or updates the site. Be sure to make note of any dates available. Other key pages to look for include Guestbook, Comments, Feedback, and Contact. These pages often have listings of comments submitted by other website visitors. In these comments

visitors may reference the claimant. For example, “Great site Christy. Thanks for emailing me with those fantastic pictures!” gives the investigator a good clue that Christy is likely the person who is actually operating the website. Guestbooks usually list the email addresses or contact information of the people who post messages, giving investigators a list of folks who may have interacted with the claimant. However, beware of relying heavily on this information as it is easily manipulated or falsified and may not be accurate. Web Logs, commonly referred to as “Blogs,” are online publicly accessible journals which may describe detailed information about a person.

### *Recovering Missing Webpages*

The Internet is ever-changing. Unfortunately for investigators who often look into the past, changes to websites occur frequently and what was posted yesterday may not be available tomorrow (Cohen, 2003, Conducting). Anyone who has encountered the familiar HTTP 404 Page Not Found! error knows how frustrating this is. Fortunately, there are ways to view the web historically.

Drill up & Drill Down. Since websites are always changing and search engine indexes are not immediately updated, sometimes the pages we expect to see are no longer available. Occasionally, this can be due to a site reorganization. In such cases, the same material may still be available but the page address or URL has changed. There are two ways to attempt to find the desired information, drilling up and drilling down. As mentioned in the Anatomy of a URL section above, the slashes in a URL indicate a subfolder and are part of the path of an html file. If that file has moved, searchers can drill up a level. To do

this, simply truncate the text to the right of the last slash in the URL (Barker, 2003, Evaluating).

It may be necessary to drill all the way up to the homepage to find a working page. As an alternative to the drill-up technique, users can drill down by going directly to the homepage and looking for new links to the information desired.

Cached Pages. Like other search engines, Google creates an index of sites that users are allowed to search. But unlike most other search engines, Google also generates a cached snapshot of the webpages in the index. That is, Google takes a picture of the page as it looks when it is indexed (Google, 2003, Cached). This snapshot is saved in the index and made available to users. The cached snapshot is updated when the site is updated in the index. Therefore, the cached version of the page is not very old. To use this feature, conduct a query using Google. When the results are listed, look for a "Cached" link under each result. Clicking this link will take you to Google's cached version of the page that is saved in the index. An extra hint: searching for a URL with Google provides some additional options.

Internet Archive Wayback Machine. Another service that provides historical views of webpages is the Internet Archive Wayback Machine. By entering a URL in the search box provided, a the user is provided with a directory of stored copies of the webpage by date. Clicking a date brings up the stored version. Sometimes graphics may be unavailable but often the page is shown intact. This tremendous resource catalogues some webpages back to 1996. While it might be an interesting novelty to see what Yahoo.com looked like back in 1996, the Internet Archive can be used as powerful investigative tool to see what a website looked like days, months, or years ago. It may also be helpful to note the progression of changes that occur on a particular website over a period of time. For example, check a

claimant's personal website for what was posted immediately before and after the disabling event.

## Searching the Web

### *Browsing*

Merriam-Webster defines browsing as “looking over or through an aggregate of things casually especially in search of something of interest.” (2003).

Browsing is a dramatically underestimated research tool. Hyperlinking encourages browsing by design. This is the essential functionality of the Internet. Investigators can browse various subjects, webring, web communities, and so on to get more acquainted with an unfamiliar topic. However, an investigator must be careful when browsing to stay focused on the task at hand. During a browsing session, a user can easily be distracted or spend several hours looking at useless or irrelevant information. Carefully controlled browsing can be effective and should not be overlooked (Weinberger, N.D.). The best investigative use for browsing is to gain familiarity with a particular web environment. For example, when locating the website of a company or individual, it may be advantageous to browse through the entire site rather than searching for a specific piece of information. Sometimes this casual looking pays big dividends when an investigator stumbles onto a previously unknown nugget of valuable information. Fortunately, users in search of more specific pieces of information can utilize the various tools available to search the Internet.

### *Search Tools*

The Internet is a vast information resource. Being able to effectively search for information is an essential tool in the Investigator’s toolbox. There are several different tools that are available for searching information on the Internet. It is important for the investigator to understand these different tools, how they work, and how to select the proper

tool for the job. After completing this chapter, refer to the Search Engine Showdown feature chart for a handy reference to the various search engine settings.

Search Engines. These tools contain indexes of the full-text of selected Web pages. They offer searching by keyword, trying to match exactly the words in the pages. Traditional search engines offer no browsing or subject categories and their databases or “indexes” are compiled by automated software programs, called spiders or crawlers, with minimal human intervention (Barker, 2003, Types). Search engines may be general or specialized according to one category of information. Pros: Targeted, full-text searching, breadth and depth of information. Cons: No annotations or browsing, not helpful for introductory information. Examples: Google, AltaVista, WiseNut, Teoma, AllTheWeb.

Meta-search Engines. These devices search multiple search engines simultaneously and return compiled results; they catch approximately 10% of search results in any of the search engines they visit. Meta-search engines can be an effective tool for searching many engines at once but usually lack the depth of results provided by more traditional search engines (Barker, 2003, Types). Pros: Semi-targeted, search multiple sources at one time. Cons: May not process search queries correctly, lacks depth of results. Examples: Dogpile, Search.com, IxQuick, SurfWax, ZapMeta.

Subject Directories. These directories include hand selected sites picked by editors, organized into hierarchical subject categories often annotated with descriptions. Users may browse subject categories or search using broad general terms. No full-text search of documents is available. Users can only search the text of the subject categories and descriptions (Barker, 2003, Types). Pros: Annotations, well organized topics, excellent for introductory information. Cons: No full-text searching, errors by human editors,

maintenance of index is questionable. Examples: Yahoo!, LookSmart, Open Directory Project, About.com.

Subject Guides. Guides are webpages containing collections of hypertext links on a subject, compiled by expert subject specialists, agencies, associations, and hobbyists. Guides are useful for getting acquainted with an unfamiliar subject or topic. They often provide links to the most popular or utilized webpages pertaining to a particular subject (Barker, 2003, Types). Frequently, these guides are denoted by the titles "Links" or "Resources." Pros: Very detailed, may have links to otherwise unknown sources. Cons: Usually no searching, questionable maintenance, reliant on a single source of input. Examples: NHCAA Resources, IASIU Links, IALEIA Links.

Specialized Databases. There is some information available through the Internet that is not searchable by the traditional search tools described above. This information resides in databases made available by various data providers (Barker, 2003, Types). These hosts provide their own search interface to this data. For more information on this topic see the section entitled "Invisible Web" below. Pros: Excellent data quality, well maintained, very targeted information. Cons: Interfaces and functionality vary, reliance on homegrown search function. Examples: LexisNexis, ChoicePoint, Accurint, ISO ClaimSearch, Ebay, Amazon.

In the early days of web searching, these different tools were easy to identify. However, some confusion among users now exists because the web search industry has undergone many changes. As a result, there are now sites that offer combined tools. Many Search Engines and Subject Directories in particular have consolidated into a one-stop search

tool. For more information about specific search tools, visit Search Engine Showdown or Search Engine Watch.

### *Search Methodology*

Investigators are trained to gather vast amounts of evidence. This mindset can cause researchers to conduct broad searches, providing millions of results. This tactic feels safe because investigators want to be sure that they are not missing some essential piece of evidence when conducting an investigation. However, with the overwhelming size of most search engine indexes, this tactic fails. Humans are simply not capable of processing the amount of information required in order to locate a specific word or phrase from billions of others. This is why search tools were designed in the first place. When conducting research on the Internet, it is best to start with a targeted query. If the desired results are not achieved, gradually expand or generalize the query.

The following five step process, developed by author James Ruotolo, is designed to help researchers locate the information they are looking for when conducting a disability insurance claim fraud investigation.

Step 1: Identify what you know. This step might seem obvious however, many researchers fail to consider what information they already have and how it might help them with their search. Remember the obvious items like name, telephone number, and address. Also consider other items like spouses, children, hobbies, or employment. If necessary, write the items down on a blank sheet of paper.

Step 2: Determine what you want to know. Identify why you are searching for information. Your search tactics for verifying information you already have will be different

than searching for new evidence. Are you trying to validate a suspicion or hunch? How specific do you need to be? Are you searching for targeted data on an individual or more general material on a particular industry?

Step 3: Select the proper search tool. Based on the previous steps, identify which search tool is best for your needs. When searching for more specific targeted information on a person or business, use a specialized database or search engine. When searching for more generic information about an industry or topic, choose a subject guide or subject directory.

Step 4: Build a query. Use the information gathered in the previous steps to build your search query. Use the terms identified in Step 1 as the basis of your query, adjusting it based on your purpose defined in Step 2. Use the proper syntax and consider default features for the tool you selected in Step 3. Be creative in generating your query and remember the GIGO principle - Garbage In, Garbage Out (Webopedia, 2003). The results are only as good as the query. Make use of Boolean operators, consider synonyms, and include name variations or aliases. See Advanced Search Techniques below for more assistance on building queries.

Step 5: Repeat. Effective web searching is an art and it takes much practice to become proficient. Repeat your search by adjusting your terms and adding or deleting items as necessary to tweak the results until you identify the information desired.

### *Common Searching Mistakes & Strategies to Avoid*

Browsing Searchable Subject Directories. Browsing can be beneficial in many ways and is an excellent method of gaining general knowledge about unfamiliar topics. However, it is not efficient to browse when looking for targeted information. Additionally, using

browsing as a search technique has other downfalls. The taxonomy in each subject directory is different and therefore classifications of topics might not be the same from directory to directory (Barker, 2002, Search). This makes it difficult to develop a consistent browsing strategy. When using a subject directory, it is best to use the search feature.

Simple Keyword Searching in Large Databases. This refers to entering vague terms in the first search box you come across as opposed to using the advanced search feature. This simple search usually utilizes the system's defaults and doesn't allow much flexibility. Simple keyword searching can be used in subject directories to guide you to the right subject area but it should be avoided in large search engines. Simple searching will result in irrelevant hits. Learn and use the advanced searching features available on search engines. Look for the "Advanced Search" link on any search engine homepage.

Focusing on Popular Links. Everyone has different information needs and interests. Simply because the site is "recommended" doesn't mean it is the best source for you to use. Recommended sites are often based on financial considerations (those sites that pay a fee to the search engine become "recommended") or link popularity (Barker, 2002, Search). Keep in mind that others might be visiting these sites for different reasons. Click carefully and make your own evaluations.

Ignoring Stop Words. "Stop Words" are words that search engines ignore because they are too common to be useful search criteria. Common stop words are adverbs, conjunctions, prepositions, and all forms of the word *be* (Sherman, 2002, Part 1). If you searched for *to be or not to be* all the words would be excluded except the word *not*. The words *to* and *be* are stop words and would automatically be ignored. The word *or* is a

Boolean operator. In this case, *not* is the only searchable word in the set. Therefore, the search engine only searched for *not*.

In this example, use of quotation marks would allow a search for the entire phrase, including the stop words. The proper search query would be: "*to be or not to be*"

It is important to note that different search engines treat stop words differently. Refer to the Help or Advanced Search features on your search site for more information.

Misusing Boolean Operators. Use of Boolean logic can strengthen a search tremendously. Unfortunately, use of Boolean operators can be confusing for some users. To make matters worse, different search engines may interpret the same operators in different ways (Sherman, 2002, Part 1). Be sure and learn how Boolean operators are used by your favorite search tools. Remember to check out what the default Boolean settings are! For more information on Boolean operators, see Advanced Search Techniques below.

Ignoring Case Sensitivity. Some search engines are case sensitive. Generally, it is best to search with all lowercase letters unless searching for a proper noun like a name or place (Sherman, 2002, Concluded). Often, this method will result in hits of all cases (UPPERCASE, lowercase, and Titlecase). Again, it is important to take note of a search engines default settings with regard to case sensitivity. To determine whether or not a search engine is case sensitive, see the search engine's Help feature or refer to the Search Engine Showdown's feature chart.

Poor Grammar. Unlike humans, computers have a difficult time determining intent. They are unable to hear inflections in tone or read body language. The only information a search engine has is the information provided in the form of search terms. Unfortunately, the overwhelming majority of novice and intermediate searchers fail to take into consideration

these subtleties of the English language. These idiosyncrasies or "Seven Deadly Nyms" (Sherman, 2002) can be the death of an otherwise good search.

Contronyms – a word that has multiple meanings that contradict the others.

Examples: Hysterical (overwhelmed with fear vs. outrageously funny). Fast (moving quickly vs. firmly stuck in place).

Heteronyms – words that are spelled identically but have different meanings when pronounced differently. Examples: bow, desert, object, lead.

Polyonyms – different words that have the same meaning. Example: Devil, Beelzebub, Lucifer, Satan.

Homonyms – words that have the same sound but a completely different meaning (and sometimes spelling). Example: to, two, too.

Capitonyms – words that change pronunciation and/or meaning when capitalized. Examples: polish vs. Polish, amber vs. Amber.

Exonym – a place name that foreigners use instead of the name that natives use. Examples: Cologne:Köln, Morocco:Morocco.

### *Advanced Search Techniques*

Once an Internet researcher moves beyond the novice stage, more flexibility in searching is often desired. All the major search engines provide an advanced search page which offers more flexibility in search logic and allows for the change of default settings used in the simple search. In addition to this functionality, a researcher can take advantage of the more advanced searching functions listed below.

Boolean Logic. Logic is used to join search data and the most common form of logic is Boolean logic. Boolean logic refers to the logical relationship among search terms and is named after mathematician George Boole (Cohen, 2003, Boolean). Almost all Internet search engines use Boolean logic as a basis for their search capabilities. Basic Boolean logic contains three search terms: AND, OR, and NOT. Many people become confused when attempting to master the process of Boolean searching because simple searches can become extremely complex and difficult to follow as additional terms are added to the search. When more terms are used with Boolean operators, the quality of search results goes up. Unfortunately, this increases the complexity of the search. For a primer in Boolean Logic, visit the University of Albany's Boolean Searching on the Internet tutorial.

Nearly all search engines on the internet allow the use of Implied Boolean Logic, also known as Search Engine Math (Sullivan, 2001). Implied search logic uses arithmetic operators instead of traditional Boolean operators (Cohen, 2003, Boolean). This helps simplify the search process for many people.

Using the OR operator. There is no arithmetic equivalent for this operator and some search engines use OR as their default Boolean setting. To search for pages that include either cats OR dogs, type cats OR dogs. Note that, in this case, OR must be capitalized to be used as a Boolean operator in some search engines.

Using the + operator. This tells the search engine that this term is absolutely required. If you were looking for pages that included both cats AND dogs, type +cats +dogs

Using the – operator. This tells the search engine to exclude a term. If you were looking for pages that include cats AND NOT dogs, type cats –dogs

Using quotation marks. Quotation marks tells the search engine to search for a phrase. Phrase searching requires all the resulting terms to appear in the same order that they were typed. If you were looking for pages that include the phrase “raining cats and dogs,” type “raining cats and dogs”

Combining operators. Combine operators to create more targeted search strings. Parentheses may also be used to execute search operators in a particular order, much like a mathematical equation. The string below searches for cats OR felines AND NOT dogs AND “grooming services.” Type +(cats OR felines) –dogs +”grooming services”

Using these Boolean operators can improve the results of your search. Remember that to use them effectively, be sure to check on how the search tool being used interprets these operators and whether or not it is case sensitive. The techniques described in this section are based on the functionality of most major search engines. However, each search engine functions differently and readers are encouraged to consult the Help feature on each search engine website before employing these techniques.

### *Power Searching*

The syntax shown for the power searching examples below is specific to Google. Search engine syntax varies and users should consult the Help or Frequently Asked Questions resources available for the search tool being used.

Title. Title Search restricts a query to the text in the HTML title of a webpage. This is the text that appears within the title tag of a webpage document. For example, if this section of this document were it’s own web page, it’s title tag might look like this <title>Power Searching</title>. To conduct a title search in most major search engines, type

allintitle:terms. To search for a web page with the term cats and the term dogs in the title, type allintitle: +cats +dogs

Site. Site search allows you to search only the pages that have been indexed for a specific website. For example, to search within Microsoft’s website for pages that include the words cats and dogs, type +cats +dogs site:microsoft.com

Domain. Site search also allows you to restrict only the domain. To search only a commercial domain for cats and dogs, type +cats +dogs site:.com. Site searching also allows you to restrict searches to country domains. Country domains are represented by a two letter designation on the end of a website URL. For example, .uk is the country domain for the United Kingdom and .ca is the country domain for Canada. To search only sites in the UK domain for the terms cats and dogs, type +cats +dogs site:.uk

Uniform Resource Locator (URL). A URL search is similar to a site search but instead searches the actual text of the URL. Users can search using the allinurl: command . To search all URLs for the term “fraud,” type allinurl:fraud

Link. The link search feature allows you to search for all the pages linking to a particular page or domain that you specify. For example, to find webpages that contain links to Microsoft.com, you would type link:microsoft.com

Wildcard. Wildcard characters, represented as an asterisk (\*), allow searching for plurals or variations of a word. It is an excellent search technique to use if the correct spelling of a word is unknown or if there are multiple spellings used. For example, to search for “theater” and “theatre,” type theat\*

Stemming. Stemming eliminates a suffix from a search term and returns all variations of that term. Some search engines provide stemming as a default. For example, if

a search is run for “singing,” results would be returned for “sing,” “singing,” “sings,” etc. In these cases, use of a wildcard symbol is not necessary.

## The Deep Web

### *Topography of the Web*

The Surface Web is the World Wide Web that most users are familiar with. It is the part of the web that consists mostly of static webpages which are indexed by traditional search engines (Sullivan, 2000, Invisible).

The Shallow Web includes dynamically generated webpages (Sullivan, 2000, Invisible). Such pages may be static pages that are dynamically delivered or may not actually exist until the user views them. For example, if a user conducts a search, the pages containing the result set are dynamically generated shallow web pages. Those pages did not exist until the user hit the search button. Search engines generally avoid indexing these types of pages as they exist for very short periods of time or their dynamic generation methods could cause the crawler software to index the same page many times accidentally (Sullivan, 2000, Invisible). They are not typically useful for investigators.

The Deep Web, which may also be referred to as the Invisible Web, is defined as web content that is not indexable because it is stored in a database. The data cannot be searched with traditional search tools like search engines (Sullivan, 2000, Invisible).

### *Navigating the Deep Web*

Deep websites tend to be much more narrowly focused on a specific topic and have a higher rate of data quality compared to the surface web (Bergman, 2001). The Deep Web holds a tremendous amount of data and is estimated to be approximately 550 times larger than the surface web (Bergman, 2001).

Even though these vast information repositories are extremely valuable, they are not generally known or utilized by most users (Bergman, 2001).

Common third-party data providers like Lexis-Nexis and ChoicePoint are obvious deep web sites that host information in a database that is accessible via the Internet but is not indexed by regular search engines. Many users may assume that deep web content is only available by subscription or other form of paid access but a recent study has found that 95% of deep web content is publicly accessible (Bergman, 2001). Some of the more common free deep web sites include Amazon.com, Ebay, the US Census Bureau, the National Oceanic and Atmospheric Association, Federal Express, and Realtor.com. All of these sites house enormous databases that cannot be searched by traditional search engines. These sites are among the most valuable resources available to investigators. Additional deep websites that are recommended for investigators are discussed in Chapter 7.

### *Searching the Deep Web*

Web surfers are accustomed to using traditional search tools to find information so the Deep Web presents a challenge for many Internet users. The best way to search deep web content has been to go directly to the website hosting the content and use that site's search function to mine the data. Various Deep Web directories have been developed to help researchers locate these valuable sites. These directories include CompletePlanet, Infomine, DirectSearch, and Invisible-Web.net.

New tools are being developed that allow users to search the Deep Web more effectively. These tools generate a meta-search by searching multiple deep web databases

simultaneously and presenting a combined result set. Profusion is an example of a Deep Web meta-search tool.

Currently there are few if any Deep Web meta-search tools that scour the free public records that are available on the Web. The best solution is to visit a Deep Web directory for public records, like SearchSystems, visit the host site, and conduct a search using the resident search functionality on that site.

## News and Newsgroups

### *Using News in Investigations*

The news information is great open source intelligence for disability insurance fraud investigators for two key reasons:

First, news entities including papers, radio programs, and television broadcast transcripts are rapidly becoming available online. There are various news aggregators that allow searching of archives and many news entities post searchable articles or archives on their own website. While some are free, many allow a free search and annotated result list, but require users to pay for the full text of the article.

Second, the fact that the information is available online may not be known to the claimant. When a person's name appears in a local newspaper, he or she may be unaware that the articles in that newspaper are posted on the Internet. Also, because people may assume limited geographic distribution of the information in a local newspaper, they may submit items like letters to the editor or place classified ads assuming only local residents will see the information, not realizing that insurance investigators have access to the same items.

Newspaper articles may have information on community groups or organizations that the claimant may be involved with. Community group activity calendars can be useful for planning surveillance or unannounced visits or interviews. Advertisements, both in the classified section and paid advertising may indicate business activity. Obituaries obviously may indicate the death of an insured. Photos or video of the claimant may be available and useful for identification purposes. Occasionally, investigators might uncover an article indicating that the claimant was involved in past insurance fraud schemes!

### *Accessing News Resources*

LexisNexis maintains an extensive full-text news archive which is searchable for a fee. NewsIndex, Newspapers.com, and ELibrary offer aggregated news archive searches. In many cases, archives are only searchable by visiting the website of an individual newspaper. A variety of tools are available to help investigators locate the daily, weekly, or monthly newspapers in various regions. These include NewsLink.org, NewsDirectory, and NewsVoyager. The Internet Public Library news section is a good resource for international newspaper websites.

For electronic media, some sites offers searches of transcripts. Transcripts.TV for example, offers the purchase of ABC News transcripts for a small fee. CNN makes recent transcripts available for free. Some radio stations or programs may have their own websites and can offer downloadable clips of segments or may be simulcast online. SpeechBot, from HP Invent, offers a free search of selected radio broadcasts, transcribed via voice recognition.

### *Newsgroup Basics*

Newsgroups are entirely different than the news resources discussed above. Newsgroups can be defined as a continuous public discussion about a certain topic (Tyson, N.D.). Sometimes referred to by the original moniker "Usenet News," they allow typed communication through an electronic medium. Unlike e-mail which is usually a one-to-one correspondence, Newsgroups are many-to-many. Essentially, Newsgroups act like electronic bulletin boards where users can post messages and all other users can read and post replies to

these messages. The Newsgroups are arranged by category and some of the common categories are Alternative (ALT), Business (BIZ), Computer (COMP), and Science (SCI).

Each of these categories is broken down into sub-categories. The names of categories are separated by periods, much like an internet address. For example, the Newsgroup containing announcements about new users of Usenet would be `news.announce.newusers`.

The Alternative category is known for its eclectic nature. Most categories require a voting process in order to create a new sub-category or group. Contrastingly, the Alternative category allows anyone to create a new group (Tyson, N.D.).

Confused about what the differences are between email, chat, newsgroups, listservs, and various other discussion mediums? [HowStuffWorks.com](http://HowStuffWorks.com) explains the differences among these forms of communication.

### *Accessing Newsgroups*

There is a Newsgroup category for practically every subject imaginable. People use Newsgroups for a variety of reasons including getting information, answers to questions, discussion of interesting or controversial topics, advertising, and so on. The attraction to Newsgroups for many users is that anonymity is possible. By setting up a free user account with a non-descript username, an individual can read and post messages without divulging his or her identity. Although some users take advantage of the opportunity for anonymity, many people do use individually identifiable information like names or addresses in their Newsgroup postings.

Historically, Newsgroups have been an area where anything goes. In other words, people are often willing to divulge information in a Newsgroup that they would not otherwise make public knowledge. Using the search feature provided by the Newsgroups, an investigator can search for any term that might appear in the text of a posting. Suggested searches might be claimant's name, claimant's email address, "insurance fraud," "disability fraud," or the name of the insurance company.

Newsgroups can be accessed via Web browser with Google Groups. Investigators are encouraged to use the search feature rather than browsing through the endless lists of postings in each of the categories. Refer to Google Groups Help which includes a basic overview of Newsgroups and a glossary of terms that are helpful to new users.

In addition to searching Newsgroups as part of an investigation, researchers can review them for valuable investigative tips and even post or exchange information. Keep in mind however, that Newsgroups are a public forum.

## Records Research

*Public Records Overview*

Public records are tremendously useful for investigators. There are many ways to obtain this type of information. Naturally, the best sources in terms of accuracy are the data collectors themselves. However, these sources are often not the most efficient or the most cost effective. Therefore, many investigators make use of acceptable substitutes in the form of third party data providers. Familiar vendors like ChoicePoint and LexisNexis provide a gateway to a goldmine of public records. Such services generally require registration and charge fees for service. However, the Internet has afforded some agencies to become more efficient at providing access to public records and many states, counties, localities, or agencies offer online access to their own data for free or for a nominal fee. Coverage and availability of such data is unpredictable and sporadic but the cost effective nature of these resources makes them worth a look. The sections below describe where to obtain information specific types of records. For a comprehensive listing of all types of public records, categorized by state, visit Search Systems.

*Telephone Numbers*

There are hundreds of white page services on the Internet. Their accuracy and freshness vary but most are useful for checking an address or the reverse search of a phone number. InfoUSA and AT&T are two of the largest data providers of free telephone and address information on the Web and their respective search services, Switchboard.com and AnyWho.com are particularly comprehensive.

In addition to white page services, other telephone tools are also available online. For example, with an area code (NPA) and prefix (NXX) free services like Fone Finder will provide basic geographical information including an area code map. In many cases these types of services will also indicate if the number corresponds to a wireless service provider. However, telephone number portability may soon eliminate the value of such tools.

For international directory listings try InfoBel.com or the Australian online white pages which includes a handy world time zone and country/province calling code utility. Although these services have many countries listed, their depth is limited and many have only yellow page listings available.

### *Addresses*

Most of the white page listing services mentioned above include address information. However, there are some additional address services that investigators might find useful. The National Address Server at the State University of New York, Buffalo allows a user to enter any mailing address and then provides the correct postal standard address for that location. The free service also provides users with a Portable Document Format (PDF) version of a mailing label with barcode that can easily be printed on an envelope or mailing label. This can be helpful when processing returned mail or ensuring that a sender has the correct postal address for any location. For complete zip code information, try the US Postal Service Zip Code Lookup utility.

For additional information on addresses, try Geocode Eagle by Tele Atlas. This service provides latitude, longitude, postal carrier, and census information about US address

locations. For maps and directions try Yahoo! Maps or MapQuest, which also offers aerial photos.

### *Property records*

Paid services generally provide the most comprehensive property records. However, many localities have provided web-based access to their property records. As an example, the city of Denver, CO offers a real estate and property tax record search tool based on address or parcel number. In New York, a non-governmental service called USPDR.com offers property searches. Unfortunately, there is no centralized group that manages these services nationwide and their availability is hit-or-miss. However, SearchSystems provides a great list of free public records available on the Web. By drilling down into their State/County/City hierarchy, users can find these types of property databases.

### *Driver Licenses & Motor Vehicle Registrations*

This type of information is very limited in availability via the Internet. Some states will allow users to check a driver license by number but will only display whether or not the license is currently valid and the expiration date. No identifying information is provided but this tool can be useful to show that the individual still has a valid driver license. For example, see Florida's Driver License Check utility.

The Drivers Privacy Protection Act, or DPPA (2002) severely restricted the states ability to offer this type of information. Specifically, in 18 U.S.C. § 2721 (a) the DPPA states that “a State department of motor vehicles, and any officer, employee, or contractor, thereof, shall not knowingly disclose or otherwise make available to any person or entity

personal information about any individual obtained by the department in connection with a motor vehicle record." Therefore, states have not provided access to this type of information via the Internet due to these dramatic restrictions. However, section (b)(6) states that it is acceptable to provide this data "For use by any insurer or insurance support organization, or by a self-insured entity, or its agents, employees, or contractors, in connection with claims investigation activities, anti-fraud activities, rating or underwriting." Therefore, driver license and motor vehicle data is available through other means – most readily through third party data providers like ChoicePoint although some states may provide the services directly for a fee.

### *Professional Licenses*

Unlike driver license information, data about professional licenses is readily available online and often for free. Many occupations have state licensing requirements: Attorneys, Public Accountants, Nurses, Hairdressers, Funeral Directors, Contractors, and Doctors for example require licensure in most states. In such cases, many states have made this license data available to the general public. For example, California allows a lookup of licensed Chiropractors. With such a tool, consumers can verify that an individual or other entity is appropriately licensed – for example, verify physician is actually a licensed doctor and not an imposter. Investigators can use this information as well. Often these databases contain detailed biographical information. A doctor's record may list school degrees, hospital affiliations, medical groups he or she is associated with, or disciplinary actions against him or her. Armed with professional license issue and expiration dates, investigators can get a sense of what activities a claimant may be involved in during a given period of time. Many

licenses require individuals to complete training on a regular basis, work a certain number of hours, or attend conferences in order to keep an active license. The issuing authority will usually list these requirements on its website. Both paid public record services and SearchSystems are great places to start searching for professional licenses.

### *Criminal Records*

Criminal history records have not found their way to the web nearly as fast as other types of records. Most vendors who allow users to request criminal records online do not deliver instant results. For the most part, the process is still manually completed. However, there is some data that is available online that may be of use to the investigator. For example, inmate records may be listed online, often with photos. The Illinois Department of Corrections Inmate search is a good example of such services. These records usually provide exact physical descriptions even if photos are not available which can be especially useful when scheduling a surveillance. The records provided on these sites are generally for informational purposes only and any information obtained should be verified through other means.

### *Business Records*

Business records may contain a wealth of information about a claimant. Involvement in a business may indicate unreported income or functionality inconsistent with the limitations of an alleged disability. While it is possible that a claimant may be working for a larger corporation, evidence of this may be hard to obtain through business records. This is better verified via tax or social security earnings records. If a claimant is involved in a small

business, perhaps a sole proprietorship, partnership, or even a small corporation, evidence of this can be found online. The reason for this is that while employee information is often scarce, data on corporate owners or officers is abundant because their names appear on corporate filings and documents. The Secretary of State (SOS) is responsible for issuing business licenses. To find the applicable SOS website, visit the National Association of Secretaries of State which includes links to each Secretary of State website. Some states offer an online corporate record lookup feature which may list officers, incorporation dates, and contact information. As always, while it is quick, easy and usually free to obtain this information over the Internet, it is best to validate the data by contacting the source directly. It can also be helpful to request copies of the actual records which often contain signatures or other identifying features not available from the web-based record.

When searching for businesses, be sure to check other sources for corporate filings as well. Most public record services offer searches for Uniform Commercial Code filings or other corporate documents. EDGAR Online offers every document filed with the SEC since 1994. Other services that have free company lookup tools include Dun & Bradstreet, Hoovers, and Business.com.

Lastly, when conducting name searches for business records, remember to search on the spouse's name as well. Often the spouse may have businesses listed in his or her name in an attempt to hide the claimant's involvement in a business.

## Organizations and Associations

### *Non-Profit Organizations*

Claimants may be involved in non-profit organizations because of the misinformed assumption that working without pay is acceptable. In cases of misrepresented functionality, such information is extremely valuable to an investigator. For information on non-profit companies and organizations, visit GuideStar or GrantSmart.

### *Sports*

Investigations of suspicious disability claims frequently turn up sports-related activity. Supposedly bed-ridden claimants have been found participating in myriad sports. Fortunately, the Internet may be able to assist investigators in discovering the true activities of a claimant. Some sporting associations or groups post member information online. For example, the Golf Handicap and Information Network allows a name and state search of members who have posted their golf handicaps online. The free search shows a golfer's home course, recent dates of play, scores, and other course information. Bowl.com offers a search for members in the ABC, WIBC, YABA, and JOG bowling leagues and the United States Pool Players Association offers a member search showing games played, dates, and home "room," or pool hall. Claim documents such as a claimant questionnaire may show what sports the claimant was involved in prior to the alleged disability. Websites for associations affiliated with these sports would be a great place to start looking. The date and location information may indicate a pattern of regular play presenting an opportunity for future surveillance.

### *Industry Trade Groups*

Claimants may be involved with trade groups or associations affiliated with a particular industry, depending on their occupation. Organization websites often list members and membership requirements. Member information may be available in a simple listing or a searchable directory format. Once membership is verified, take note of membership requirements. Often, members are required to attend annual meetings, receive a number of hours of training, or maintain some type of related employment. It is also helpful to take note of any major activities that the group may be planning. Seminars, in-service training, or even merchandise may be associated with the claimant. For example, an engineer who is on disability may begin lecturing at training seminars or publishing materials sponsored by a trade association. For examples of trade associations, visit the Yahoo! Trade Associations Directory. To search for a particular type of association, visit the American Society of Association Executives search utility.

### *Hobbies*

In addition to sports and employment-related activities, the claimant may be involved in other social organizations. Review the claim file and identify any other hobbies or activities that the claimant has listed on forms or mentioned in conversation. Consider researching membership in these groups: Local government, committees, civic organizations, fan clubs, volunteer programs, church or religious groups, collectors groups, and youth assistance programs.

Many of these groups have websites. Often information about current or upcoming events is listed and occasionally photos, and increasingly videos, from group events are posted.

## Anonymous Investigation

### *Electronic Footprints*

The element of surprise is essential in most investigations. What many novice Internet users fail to realize is that visiting a website allows the owner of the website to view certain characteristics about you that often divulge your identity. To keep your identity secret, it is important to engage in safe surfing practices.

To see what others can learn about you just from your visit to a website, obtain a free analysis from Privacy.net. With this information obtained, anyone can look up your registration information (Smith, 1998). If your internal network is protected by a properly configured firewall, your individual computer's Internet Protocol (IP) address will not be visible to the Web. However, your corporate network's IP address may be visible. With that information, a savvy user can identify which company you work for and maybe where you are physically located. For a more comprehensive review of IP addressing, see *Understanding IP Addressing: Everything You Ever Wanted to Know* (3Com Corporation, 2001).

### *Cookies*

Cookies are small text files saved on a local computer that contain information generated by websites based on user interaction (Webopedia, 2003, Cookie). Some websites use them to enhance the browsing experience and many sites do not work properly unless a browser is configured to accept them. Cookies are incredibly useful and allow websites to confirm and remember a user's identity, preference, and past browsing history. This makes targeted delivery of web content possible and is a great feature of many websites. For

example, if a user enters a zip code on a weather website, the user will see a local weather report every time they visit that site if a cookie stores the zip code information.

Cookies can however, be used for less desirable purposes and may be a privacy concern (Privacy.net, N.D., Bake). Since they store information about a users surfing preferences, activity, and possibly more sensitive data, it is important to actively manage cookies. Most browsers have a default setting to accept all cookies. Cookies can be managed in one of two ways: with third party software or via the Internet browser. Third party software can be purchased to manage cookies and many common security utility software packages include cookie-management utilities. This option provides the most flexibility. Alternatively, to manage cookies via a browser, change the browser's options or preferences to not accept cookies or to ask permission before accepting a cookie.

### *Online Anonymity*

There are several ways to cloak your identity when surfing online. One method is to obtain Internet access through an Internet Service Provider, independent of your employer. However, it is not a good idea to trace the Internet service plan back to you as an individual. Law enforcement entities can setup fake identities for the sole purpose of obtaining independent Internet service. Private sector investigators do not have that luxury.

Other solutions involve masking a real Internet account and IP address. The simplest method is to utilize a proxy service. Web-based proxies allow users to enter a destination Web address and then filter the data through their service so that the target sees only the anonymizer service information and not the identity of the user. There are advantages to using these types of services: some are free, all are available online and do not require

installation of any software so they generally work regardless of your computer system configuration. The free versions result in some performance degradation and many offer additional pay services for full bandwidth. Since most proxies encrypt your browsing sessions, they allow users to skirt the firewall filtering many companies put in place to keep employees from viewing unauthorized material from their work computer. Because of this, many companies filter the sites used to access proxies. Contact a representative from your information technology (IT) department to assist with this type of service.

Lastly, there are software packages available for purchase that provide the same functionality. These must be installed on a local computer or network to operate and need to be configured appropriately. In the private sector environment, this solution usually requires buy-in from purchasing and procurement, IT services and Information Security departments and, as a result can be more difficult to implement.

#### *Anonymous Email Communications*

Email communications are also traceable. The header information contained in an email provides detail about where the message originated and the path it took to arrive at the recipient's inbox (Privacy.net, N.D., Being Traced). Anonymous email remailers exist which can be used to send anonymous email to anyone. For a sample list of these services, check the Google Directory. Anonymous email can be useful for investigators to send anonymous messages to others. However, the problem with this method is that it is only effective as one-way communication. Anonymous email does not allow the recipient to reply to the originator. There are some exceptions where a third party remailer will act as an intermediary between two individual email accounts. This method is frequently used by

personal ad and dating websites to protect the identity of members until both communicating parties agree to exchange identifying information. If repeat or reply communication is needed, it is preferable to setup an email account with an undercover persona.

### *Establishing an Undercover Online Persona*

By establishing an online identity, an investigator can effectively communicate and interact with individuals, businesses, and web services without compromising the element of surprise. When utilizing an undercover identity in the online world, it is best to do so via the use of an anonymizer service, thus masking the investigators true location. Setting up an identity is a four step process: Establish a persona, obtain an email account, sign up for other services, and manage the identity.

Establish a Persona. Depending on what information is needed and how the identity will be used, it is necessary to prepare some information before embarking on the remaining steps in the process. Generally, the following information will be necessary to proceed: Name and gender; city, state, and zip code; date of birth and age; and occupation.

Consider how the identity will be used. If there will be direct communication with a target or suspect, consider how that individual will respond to a male versus female identity. Choose a name carefully. Common names like John Smith are not effective as they are too obviously vague. Likewise, extremely unusual names can draw unwanted requests for additional personal information, like inquiries about nationality or personal family history. Make sure the city, state, and zip code match because some registration applications verify that the city and zip code refer to the same location. A physical address is generally not necessary for most scenarios but it can be helpful to be familiar with the area chosen in case

discussions about local geography develop. Carefully consider the date of birth to ensure that the age is appropriate for the identity. Also consider how others who will be recipients of communications from the identity will respond to an individual of a given age. Make sure that the age matches the date of birth correctly, taking into account any leap year complexities. Take note of any holidays which fall on the date that is chosen. It is best to avoid choosing major holidays as birthdates as these will likely result in more unwanted questions. Also, remember to keep track of the age of your identity, it changes every year! Choose an occupation that is appropriate for the purpose and consistent with the other aspects of the identity. Making a 16 year old girl the CEO of a Fortune 500 company is not recommended. Avoid company names unless absolutely necessary. Often, only a vague occupational description is required for registration with “Consultant” or “Student” among favorite choices. Avoid titles which indicate extensive knowledge in a subject area, especially if you are not knowledgeable. While “Software Developer” is sufficiently vague, “Visual Basic Programmer” implies specific knowledge. Also avoid titles that require or often are associated with particular certifications. For example, a “Database Administrator” would likely have a vendor certification of some type.

After this information is established and recorded for future reference, some thought should be given to other aspects of the identity. Examples of items to consider are: Education, computer skill level, online activity, and hobbies or other interests.

Once decisions are made regarding these characteristics, they should be recorded and should not change dramatically. Investigators can allow a realistic progression over time if they wish although it is common to use an identity for only a short time.

Obtain an Email Account. The first web stop that this new “person” needs to make is a free email service provider. Hotmail or Yahoo! Mail are common choices. When setting up a free email account, carefully choose an account ID as this will be used in the email address itself. Account IDs and associated email addresses must be unique. Common configurations like FirstnameLastname (e.g. JohnSmith) are usually unavailable because they have already been secured by another user. Consider using numbers (e.g. Johnnie42) or alternative spellings (e.g. HockeePlaaer) to produce a unique ID. The email account is the basis for the identity. Cautiously choose whether or not to have the email account listed in the directory, the email equivalent of a phone book. Once the address has been created, login to verify that the address works, check any new user or welcome messages which may contain details about account usage and adjust account options to the preferred settings. Take note that using a free email service does not provide total anonymity. Email headers even in mail generated by these free services may contain information linking you to your Internet Service Provider (Privacy.net, N.D., Being Traced). It is best to use these accounts in combination with some form of proxy software.

Signup For Other Services. With an active email account, a user can sign up for myriad other services. Depending on what the needs of the investigation are, it is recommended that the investigator subscribe to many of the common Internet services. Often premium access is afforded only to members. For example, with many services like America Online or Ebay, searching for a member profile is allowed only for active members. Therefore it is important to establish membership with the covert identity and email address previously created. IDs and passwords need to be chosen according to the same parameters

described above for an email account. Services that an investigator may wish to register for include: Ebay, Yahoo! Groups, AOL Instant Messenger (AIM), and ICQ.

It is common for sites that require registration to send a confirmation email to your active email account, providing you with an additional one-time access code that needs to be used for your first sign-in. This ensures that the service has obtained a valid email address from the user and is the reason that investigators should setup an email account first. Check the email account soon after registering with these other services to see if any action is required to complete the registration process. Then login to the services and adjust the user options to the preferred settings. Become familiar with the layout and functionality of the service, member profile system, and most importantly, search options. Use access to view other member profiles, access account histories or transact and communicate with other users as part of an investigation.

Manage the Identity. After setting up the accounts, the investigator needs to manage them. The most essential requirement is to regularly check all email accounts associated with the covert identity that has been created. This is necessary for two reasons. First, unsolicited commercial email, or spam floods inboxes. Most free email providers like those mentioned above have mailbox size limitations. Once the mailbox is full, new messages will no longer be accepted. It is important to empty the mailbox from time to time to ensure that desired communications are able to be received. Second, people may send mail to the address. Though an investigator may not be expecting it, a target or subject of an investigation may send an email. The email address may have made its way to others who could be sending important messages. To keep up the authenticity of the identity it is important to maintain regular communication.

Keep in mind that the identity being used should age appropriately over time. Changes in employment, marital status, and even hobbies should be tracked. Each investigation has a different set of circumstances and requires varying degrees of involvement. One preferred method is to keep one identity for lookup purposes only. Use this identity only to view profiles or gather information but not to communicate or interact in any way. Use other identities for email, chat, messaging, newsgroups, listservs, and so on. Finally, remember to retire an identity after it has served its purpose. While it is possible to reuse accounts and IDs several times for separate investigations, reusing accounts for separate purposes can have unintended consequences. Remember that geographic distance is not a reliable reason to assume that two people do not interact with each other online. Most accounts including email will become inactive and then terminate after a certain period of time. Be aware of these thresholds to avoid premature termination.

**Important Note:** It is extremely important that users read and fully understand the terms of use and subscriber agreements associated with all services. Investigators should take care to ensure that their actions do not violate the terms of these agreements. The author does not condone usage of any service outside the terms of the subscriber agreement.

References

- 3Com Corporation. (2001). Understanding IP Addressing: Everything You Ever Wanted To Know. Retrieved November 25, 2003 from [http://www.3com.com/other/pdfs/infra/corpinfo/en\\_US/501302.pdf](http://www.3com.com/other/pdfs/infra/corpinfo/en_US/501302.pdf)
- Barker, J. (2003). Evaluating Web pages: Techniques to apply & questions to ask. Finding Information on the Internet: A Tutorial. University of California at Berkeley. Retrieved November 23, 2003 from <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Evaluate.html>
- Barker, J. (2003). Glossary of Internet & Web jargon. Finding Information on the Internet: A Tutorial. University of California at Berkeley. Retrieved November 23, 2003 from <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html>
- Barker, J. (2003). Search strategies: Search with peripheral vision. Finding Information on the Internet: A Tutorial. University of California at Berkeley. Retrieved November 23, 2003 from <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Strategies.html>
- Barker, J. (2003). Types of search tools. Finding Information on the Internet: A Tutorial. University of California at Berkeley. Retrieved November 23, 2003 from <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/ToolsTables.html>
- Bell, C. (2003). Anatomy of a URL. Web Publishing Curriculum Resources. University of Oregon Libraries. Retrieved November 23, 2003 from <http://libweb.uoregon.edu/it/webpub/url-anatomy.html>

Bergman, M. K. (2001). The deep web: Surfacing hidden value. *Journal of Electronic Publishing*. Volume 7, Issue 1. University of Michigan Press.

<http://www.press.umich.edu/jep/07-01/bergman.html>

Christensen, D. (2003, June 24). Driven to sue: Suits in West Palm Beach allege personal information on state's 13 million drivers being sold unlawfully [Section: From the courts]. *Broward Daily Business Review*, 3. Retrieved August 12, 2003 from LexisNexis.

Cohen, L. (2003). Conducting research on the Internet. Internet Tutorials. University at Albany Libraries. Retrieved June 22, 2003 from

<http://library.albany.edu/internet/research.html>

Cohen, L. (September, 2003). Boolean searching on the Internet. *Internet Tutorials*..

Retrieved June 22, 2003 from University at Albany Libraries:

<http://library.albany.edu/internet/boolean.html>

Conning & Company. (2000). *Insurance Fraud: Renewing the Crusade*.

Delizia, Michael. (2003). Glossary of Disability Insurance Terms. *Disability-Expert.com*.

Retrieved June 22, 2003 from <http://www.disability-expert.com/glossary.htm>

Dictionary.com. (n.d.). What is the difference between data and information? *Frequently*

*Asked Questions*. Retrieved November 23, 2003 from

<http://dictionary.reference.com/help/faq/language/d/data-info.html>

Driver's Privacy Protection Act, 18 U.S.C. § 2721 (2002).

Fair Credit Reporting Act, 15 U.S.C. § 1681 (2002).

Google. (2003). Cached links. *Google Web Search Features*. Retrieved November 23, 2003

from <http://www.google.com/help/features.html>

- Individual Reference Services Group. (1997). IRSG Principles. *Individual Reference Services: A Report to Congress*. Appendix D. Retrieved November 23, 2003 from the Federal Trade Commission website: <http://www.ftc.gov/os/1997/12/irsappd.pdf>
- Jesdanun, A. (2003, November 2). Internet littered with dead web sites. *Miami Herald*. Associated Press. Retrieved November 24, 2003 from <http://www.miami.com/mld/miamiherald/business/7165954.htm>
- Kirk, E. E. (2002). *Evaluating information found on the Internet*. Retrieved November 23, 2003 from the Sheridan Libraries of The Johns Hopkins University website: <http://www.library.jhu.edu/elp/useit/evaluate>
- McRae, J. (September 2001). The evolution of insurance fraud detection: Lessons learned from other industries. *Claims Magazine*. Retrieved August 12, 2003 from <http://www.claimsmag.com/Issues/Sept01/evolution.asp>
- Merriam-Webster. (2003). Browse. *OnLine Dictionary*. Retrieved November 23, 2003 from <http://www.m-w.com/cgi-bin/dictionary>
- Privacy.net. (N.D.). *Bake Your Own Internet Cookie!* Retrieved November 25, 2003 from <http://www.privacy.net/cookies>
- Privacy.net. (N.D.). *Being Traced Over the Internet*. Retrieved November 25, 2003 from <http://www.privacy.net/Traced>
- Sherman, C. (2002, March 27). Seven stupid searching mistakes, part 1. *SearchDay*. Retrieved November 23, 2003 from <http://www.searchenginewatch.com/searchday/article.php/2159561>

Sherman, C. (2002, March 28). Seven stupid searching mistakes, concluded. *SearchDay*.

Retrieved November 23, 2003 from

<http://www.searchenginewatch.com/searchday/article.php/2159571>

Sherman, C. (2002, August 7). The seven deadly nymys. *SearchDay*. Retrieved November

23, 2003 from <http://www.searchenginewatch.com/searchday/article.php/2160501>

Smith, R. (1998). The IP address: Your Internet identity. *Consumer.net*. Retrieved

November 25, 2003 from <http://consumer.net/IPpaper.asp>

Sullivan, D. (2000, August 2). Invisible web gets deeper. *SearchDay*. Retrieved November

23, 2003 from <http://searchenginewatch.com/sereport/article.php/2162871>

Sullivan, D. (2001). *Search Engine Math*. Retrieved November 23, 2003 from

<http://www.searchenginewatch.com/facts/article.php/2156021>

Tyson, J. (n.d.). How newsgroups work. *HowStuffWorks*. Retrieved November 24, 2003

from <http://computer.howstuffworks.com/newsgroup.htm>

Webopedia. (2003). *Cookie*. Retrieved November 25, 2003 from

<http://www.webopedia.com/TERM/c/cookie.html>

Webopedia. (2003). *DNS*. Retrieved November 26, 2003

<http://www.webopedia.com/TERM/D/DNS.html>

Webopedia. (2003). *garbage in, garbage out*. Retrieved November 23, 2003 from

[http://www.webopedia.com/TERM/g/garbage\\_in\\_garbage\\_out.html](http://www.webopedia.com/TERM/g/garbage_in_garbage_out.html)

Webopedia. (2003). *TLD*. Retrieved November 26, 2003 from

<http://www.webopedia.com/TERM/T/TLD.html>

Webopedia. (2003). *Internet*. Retrieved June 22, 2003 from <http://www.webopedia.com/>

[TERM/I/Internet.html](http://www.webopedia.com/TERM/I/Internet.html)

Webopedia. (2003). *World wide web*. Retrieved June 22, 2003 from

[http://www.webopedia.com/TERM/W/World\\_Wide\\_Web.html](http://www.webopedia.com/TERM/W/World_Wide_Web.html)

Weinberger, D. (N.D.). Fear of browsing. Intranet Journal. Buzz Soup. Retrieved

November 23, 2003 from <http://www.intranetjournal.com/opinion/fear.shtml>

WordNet. (2003). Overview for insurance claim. WordNet 1.7.1 Search. Cognitive Science

Laboratory. Princeton University. Retrieved June 22, 2003 from

<http://www.cogsci.princeton.edu/cgi-bin/webwn1.7.1?stage=1&word=>

[insurance+claim](http://www.cogsci.princeton.edu/cgi-bin/webwn1.7.1?stage=1&word=insurance+claim)