INTELLIGENCE EXPLOITATION OF THE INTERNET

OCTOBER 2002

TABLE OF CONTENTS

INTRODUCTION

This handbook is the third in a series of publications produced by SACLANT to improve the understanding of Open Source Intelligence (OSINT) within NATO. The first publication, the NATO OSINT Handbook served as an introduction to the field of OSINT. The second volume, the NATO OSINT Reader was intended to provide under one cover a collection in international writings on the uses of open information sources in the preparation of intelligence products.

This volume is entitled Intelligence Exploitation of the Internet. In contrast to the first two volumes, this publication is intended primarily as a practical guide to the exploitation of an information source. While many people wrongly equate the Internet with the totality of open sources, it does currently serve as an important information source for many intelligence staffs. Although much information is available on the Internet, the Internet is in reality a communications medium upon which information flows rather than an information repository in its own right.

The Internet has largely replaced dedicated stand-alone communications networks used by commercial information providers to distribute their products. Specialized information such as ship and aircraft movement data, along with virtually all commercially published material can be accessed via the Internet through commercial information vendors. These sources often provide the most valuable unclassified intelligence data. The use of data distribution services such as FACTIVA, LEXIS-NEXUS or DIALOG requires specialized training to match sources to information requirements and thus remains outside the scope of this publication.

Rather, the aim of this publication is to expose intelligence staffs to many of the challenges of using information found via the Internet for intelligence purposes. It is structured according to the intelligence cycle.

CHAPTER II: DIRECTION underscores the importance of ensuring that information needs stem from mission requirements. The process of conducting mission analysis leading to the preparation of Primary Intelligence Requirements is explained.

CHAPTER III: COLLECTION begins with the process of preparing a collection plan using Internet sources. It continues with an examination of search strategies and tools. It concludes with a short discussion of the dangers associated with exposing intelligence interests through the trail that an Internet search leaves on the Internet.

CHAPTER IV: PROCESSING focuses on how to shape information gathered from the Internet, adding analysis, into intelligence – information that is relevant, reliable and that contributes to the generation of knowledge. It begins with how to analyse the source of the information. While classified intelligence collection disciplines typically discriminate between evaluated and unevaluated information, all data gathered from the Internet must initially be treated as unevaluated. It is through the process of source evaluation by an

analyst that information can be given greater credence. The Processing Chapter concludes with descriptions of tools that aid in the maintenance of validated sources and the importance of reducing information overload.

This volume concludes with CHAPTER V: DISSEMINATION. This chapter outlines a number of dissemination advantages that arise through the production of intelligence from unclassified sources. In particular, the advantages of using open sources to support the production of intelligence for coalition operations are explained.

The annexes include additional amplifying material. Annex A: More information on Source Evaluation provides some additional techniques to use in the evaluation of information sources prior to accepting information contained within their web-pages. Annex B: Selected Information Sources was developed by CINCEASTLANT Intelligence Staff and contains practical search tips, techniques and sources for Internet research on a number of relevant NATO intelligence topics of interest.

Much of the information contained in this volume has been copied directly from Internet sources. Rather than try to replicate many of the excellent training sources already in existence, this volume gathers together a number of relevant published writings. The source has been cited wherever possible. Care has been taken to organize the resulting collection into a user-friendly format. Additional topics have been addressed with original material so as to ensure that this volume comprises a complete introduction to intelligence exploitation of the Internet.

CHAPTER I: INTERNET OVERVIEW

Section A: World-Wide-Web (WWW)

1. What is the Internet?

The Internet is a network of networks, linking computers to computers sharing the TCP/IP protocols[1]. Each runs software to provide or "serve" information and/or to access and view information. The Internet is the transport vehicle for the information stored in files or documents on another computer. It can be compared to an international communications utility servicing computers. It is sometimes compared to a giant international plumbing system. The Internet itself does not contain information. It is a slight misstatement to say a "document was found *on* the Internet." It would be more correct to say it was found *through* or *using* the Internet. What it was found in (or on) is one of the computers linked to the Internet.

Computers on the Internet may use one or all of the following Internet services:

- Electronic mail (e-mail). Permits you to send and receive mail. Provides access to discussion groups often called Listservs® after the software they operate under.

- Telnet or remote login. Permits your computer to log onto another computer and use it as if you were there.

- FTP or File Transfer Protocol. Allows your computer to rapidly retrieve complex files intact from a remote computer and view or save them on your computer.

- Gopher. An early, text-only method for accessing Internet documents. Gopher has been almost entirely subsumed in the World Wide Web, but you may still find gopher documents linked to web pages.

- The world wide web (WWW or "the web"). The largest, fastest growing activity on the Internet.

2. What is the World Wide Web and what makes it work?

The WWW incorporates all of the Internet services above and much more. You can retrieve documents, view images, animation, video, listen to sound files, speak and hear voice, and view programs that run on practically any software in the world, providing your computer has the hardware and software to do these things.

---

[1] TCP/IP (Transmission Control Protocol/Internet Protocol) is the basic communication language or protocol of the Internet. It can also be used as a communications protocol in a private network (either an intranet or an extranet). When you are set up with direct access to the Internet, your computer is provided with a copy of the TCP/IP program just as every other computer that you may send messages to or get information from also has a copy of TCP/IP.

When you log onto the Internet using Netscape or Microsoft's Internet Explorer or some other browser, you are viewing documents on the World Wide Web. The current foundation on which the WWW is based is the programming language called HTML. It is HTML and other programming imbedded within HTML that make possible Hypertext. Hypertext is the ability to have web pages containing links, which are areas in a page or buttons or graphics on which you can click your mouse button to retrieve another document into your computer. This "clickability" using Hypertext links is the feature that is unique and revolutionary about the web.

How do Hypertext links work? Every document, file, site, movie, sound-file or anything you find on the web has a unique URL (Uniform Resource Locator) that identifies what computer the item is on, where it is within that computer, and its specific file name. Every Hypertext link on every web page in the world contains one of the URLs. When you click on a link of any kind on a web page, you send a request to retrieve the unique document on some computer in the world that is uniquely identified by that URL. URLs are like addresses of web pages. A whole cluster of internationally accepted standards (such as TCP/IP and HTML) make possible this global information retrieval phenomenon that transcends all political and language boundaries.

3. What is a Browser?

A browser is a program that resides on your computer enabling you to view WWW documents and access the Internet taking advantage of text formatting, hypertext links, images, sounds, motion, and other features. Netscape and Internet Explorer are currently the leading "graphical browsers" in the world (meaning they facilitate the viewing of graphics such as images, video and more). There are other browsers (e.g., Macweb, Opera). Most offer many of the same features and can be successfully used to retrieve documents and activate many kinds of programs.

Browsers all rely on "plug-ins" to handle the "fancier" files you find on the web. Plug-ins are sub-programs stored within a browser or elsewhere in your computer especially to support special types of files you may click on. If you click on a link, and your computer does not currently have the plug-in needed for the file you clicked on, you are usually prompted with an opportunity to get the plug-in. Most plug-ins are free, and easy and safe to install on your computer; follow the instructions you are given.

The main way in which browsers differ is in the convenience features they offer for navigating and managing the web and all the URLs you may want to keep track of. Netscape and Internet Explorer both offer the ability to e-mail documents, download them to diskette, print them, and keep track of where you've been and sites you want to "bookmark."

Section B: Newsgroups

There are over 60,000 Usenet newsgroups where Internet users exchange messages on specific subjects. If you can't find something online via search tools, then try the newsgroups. A fellow Internet user can probably lead you to the right information.

Main points:

- Usenet Newsgroups are a globally shared/distributed set of online forums where users can post messages to each other.

- Newsgroups can be subject oriented (alt.news.macedonia) or created based on geography (soc.culture.afghanistan).

- Newsgroups are not chat rooms. Chat rooms are a more real-time event, where the text of messages is displayed back and forth between users in near real time. Newsgroup postings are more similar to email messages that are posted into a shared area, where other users may reply back to the message over the next several days.

- Quality of content in a specific newsgroup depends on the culture of that newsgroups' following. This is an online city of over 100 million strangers.

- Take the time to watch and learn the culture of a group before participating (called lurking).

- Thou Shall **not** SPAM[2] the Internet!!!

- Recommended Newsgroup reader software is Forte Agent.

---

2 What is SPAM?
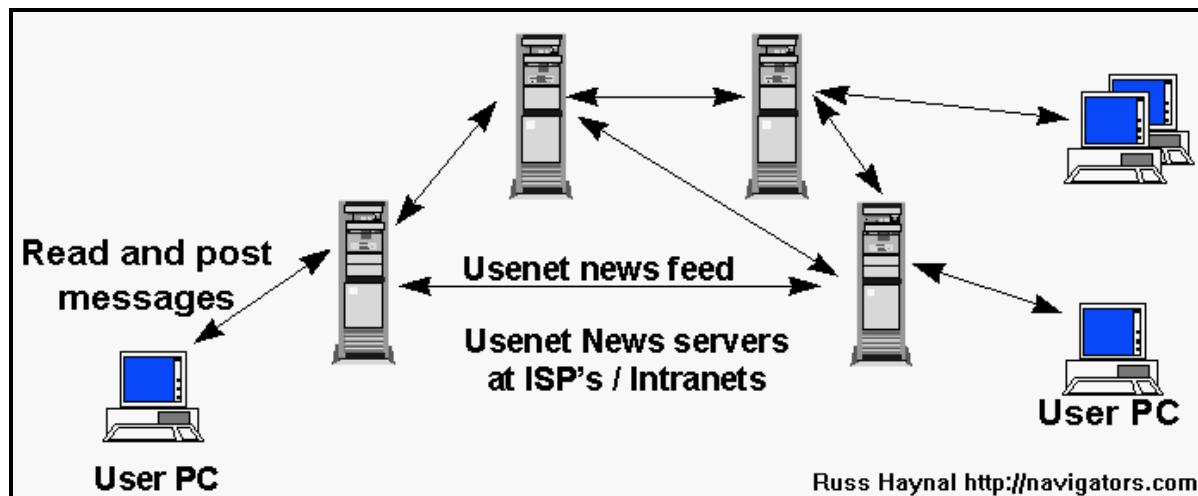
Spam is flooding the Internet with many copies of the same message, in an attempt to force the message on people who would not otherwise choose to receive it. Most spam is commercial advertising, often for dubious products, get-rich-quick schemes, or quasi-legal services. Spam costs the sender very little to send -- most of the costs are paid for by the recipient or the carriers rather than by the sender.

	• Messages (articles) are "posted" to the newsgroup
misc.invest.stocks Free SEC Filings?? My broker shorted me! FAQ - All about investing	• Others may post their "follow-up" to the newsgroup OR may respond directly to the author via email • A series of related "articles" is called a "thread" • A FAQ (Frequently Asked Questions) may be posted. READ IT, to understand the purpose of the group, and to see if "your" question" is already answered in the FAQ. A list of FAQs is also available via the web.

Usenet Newsgroup Architecture

Read and post messages

Usenet news feed

Usenet News servers at ISP's / Intranets

User PC

User PC

User PC

Russ Haynal http://navigators.com

- Newsgroups are hosted on news servers throughout the Internet.

  o News Server onsite (i.e. news.company.com)

- o Use provider's server (i.e. news.ISP.net)

- Your local Newsreader is used to interact with this information. This can be a stand-alone Newsreader, or your web browser may also provide this capability. Check with your local Internet provider or network administrator for the software settings.

- You read and post messages to a specific new server. New postings are then fed to other new servers throughout the Internet, until each news server has a roughly equivalent set of messages.

- There are also public web-based news servers such as Dejanews which can be used by anyone.

Searching: There are several places on the Internet where the content of the newsgroups are being archived and made available for full-text searching: Dejanews, remarq, Altavista Usenet search, Liszt, tile.net. Google Groups.

For More Information:

- Usenet Topology -shows a small portion of usenet topology (in Italy)

- Top 1000 Usenet Sites - Shows which news servers are widely propagated.

- Emily Post News - great examples of what **not** to do in a newsgroups - a must read.

- Usenet death Penalty - See how the community can go after ISP's who permit Spam Abuse.

- Free advice on Usenet news - a fairly complete overview.

- NNQLINKS - news.newuser.questions contains a useful collection of information.

- Creating New Newsgroup - explains the process very well.

- Usenet Newsfeeds via satellite: Cidera, ispsat.

Source: http://navigators.com/usenet.html

Section C: Email List Overview

There are over 100,000 email lists where communities of Internet users share information and advice with each other.
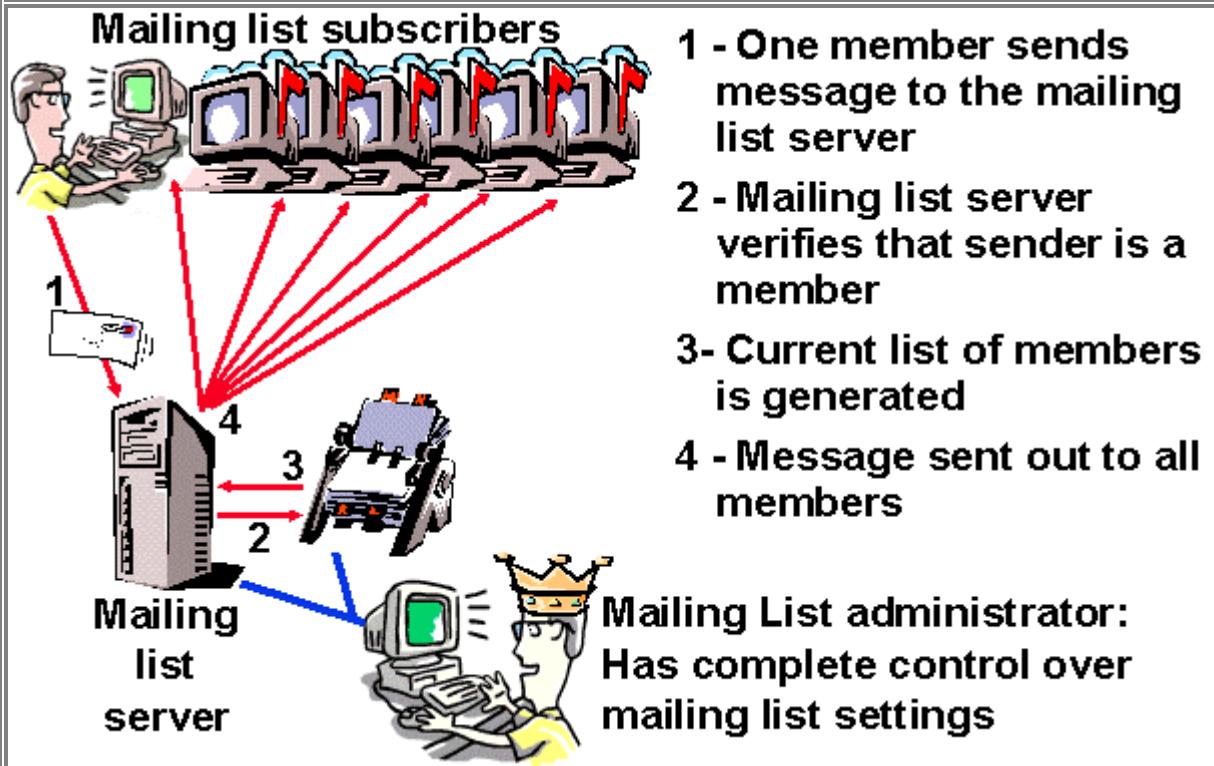
Main points:

- A mailing list provides an easy way for a large group of people to have an ongoing discussion via email.

- Most mailing lists are subject oriented.

- "Someone" must establish the mailing list and have it "hosted" at some list server. The "creator" of the mailing list is usually referred to as the list administrator.

- Mailing list software is often used to automate many of the list's functions such as joining a list and relaying emails to all list members.

- The mailing list will have an email address such as topic_name@hosting-site.com.

- Users must then **subscribe** to the list in order to receive the content of the mailing list. Subscribing is often done via an email command sent to the list_server [see above] software. For example:

> To: listserv @ hosting-site.com
> from: your_email @ your-site.com
> subject: (you can leave this blank)
>
> subscribe topic-name

- As you subscribe, the very first message you receive will probably be a canned welcome message that the list_server send to all new subscribers. **Save** this welcome message. This message tells you what is the purpose of the list, what's allowed to be posted to the list, and most importantly, how to unsubscribe when you want to leave.

- At this point you might also begin receiving a flood of frequent email depending on the volume of the email list. You should definitely make a filter rule in **your** email client to handle all this new email.

- It is suggested that you "listen to the list for a while, before you jump in and participate (this is called "lurking").

- Some lists will allow any member to send a message to the rest of the group (Discussion-style) On other mailing lists, the mail administrator may be the only person allowed to post messages, much like a newsletter broadcast (announcement-style).

Sending an email to a mailing list

Mailing list subscribers

1 - One member sends message to the mailing list server

2 - Mailing list server verifies that sender is a member

3 - Current list of members is generated

4 - Message sent out to all members

Mailing list server

Mailing List administrator: Has complete control over mailing list settings

- Businesses can use mailing lists as a way to make announcements, send newsletters, or host product discussion groups.

- Mailing lists are especially useful for timely information and distributing information to a large number of people.

- Mailing lists can have some of the best information about a topic, because all the participants have made a commitment to receive all the messages being posted to the list.   Unlike the newsgroups, something can be done about individuals who are abusing the groups' discussion. If enough people complain, the all-powerful list administrator has the ability to remove any individual who is not welcome on the list anymore.

- Searching for a mailing list? try PAML, CATALIST, TOPICA   Meta-list.

For more information see:

- Email list software vendors: listserv, majordomo.

- Listserv top 20 - A daily indication of the largest mailing Listservs.

- Internet mailing list providers - this is useful for anyone who would like to create an email list.

Source: http://navigators.com/email_lists.html

Section D: Chat

Unlike the other communications venues described so far, chat rooms are synchronous. The conversations occur in real time, just like a conference telephone call. Chat rooms are typically found on specially programmed web pages. You type a message on the chat room screen and all the other people in the chat room see your name (or 'handle') followed by the message you are typing.

The best thing about chat rooms is that the communication back and forth is immediate. The conversation moves along as fast as the participants can read the messages and type in their replies. The disadvantages are that having more than 5-10 people in the chat room makes it difficult to follow the conversation because new messages are appearing on the screen before you have finished typing in your reply to a previous one. When your reply shows up on everyone else's screen, they may not know exactly what message your reply referred to.

The fact that everyone has to be there at the same time is also a disadvantage. The use of Chat Rooms for intelligence purposes begins to cross the threshold between OSINT (primarily passive) and HUMINT (primarily active collection).

CHAPTER II: DIRECTION

Section A: Mission Analysis

The classified intelligence production process within virtually all NATO Commands and member countries provides a stream of intelligence products. These are tailored to the intelligence needs of the recipient staffs based on established requirements. Access to a robust collection of open sources enables intelligence staffs to gather and process relevant information to supplement classified intelligence products. However, it provides much extraneous information as well.

It is folly to attempt to replicate on classified systems the host of relevant, reliable knowledge available from open sources. It is equally unwise to merely attempt to find information that may be relevant to a decision-maker on a topic that appears of interest. All efforts to effectively exploit any open source of information must be guided by mission requirements. Relevance is the key. Effective open source exploitation requires appropriate selection of sources and information.

"Relevance" judgments typically stem from the commander's Priority Intelligence Requirements (PIRs), which outline that intelligence deemed necessary by him to accomplish his mission. While ideally intelligence requirements should come directly from the commander, in reality, it is up to his intelligence staff to draft them based on an understanding of his intelligence needs. This process begins with mission analysis.

Within a staff structure, each staff element responsible to support the planning function with the development of estimates begins their work with their own mission analysis. As the intelligence staff is expected to provide the foundation knowledge upon which other planning is to be conducted, effective mission analysis is essential to generate appropriate intelligence products. Once mission analysis is complete, PIRs can be developed to support the mission needs. These PIRs in turn form the outline of collection plans where open sources are able to contribute alongside other intelligence sources.

Steps in Mission Analysis

Mission analysis is a part of the problem-solving technique that military staffs use to study a mission and to identify all tasks necessary to accomplish the mission. The process is as applicable to developing PIRs in advance of a military operation as it is in developing an intelligence production plan to support situational awareness. Mission analysis produces an understanding of the commander's information requirements based on an appreciation of the commander's responsibilities and those of his superior commander. Mission analysis is the primary factor in the development of an estimate. It

is also the initial step in understanding the relevance of an issue to a particular decision-maker (e.g. The threat of a South Asian war to a NATO commander) and those component elements of the issue that affect his mission (e.g. Danger of escalation/Impact on allied operations in theatre/etc).

When a commander receives a mission tasking, normally as a Warning Order, analysis begins with the following questions:

- What tasks must my command do for the mission to be accomplished?
- What is the purpose of the mission received?
- What limitations have been placed on my own forces' actions?

Once these questions have been answered, the commander can thoroughly understand the mission. When conducting OSINT exploitation in support of situational awareness, a similar set of questions emerge:

- How is this event relevant to my commander?
- What impact does this event have on assigned forces or potential operations?
- Will this event shift the operating environment within his Area of Operations?

Mission analysis steps relevant to the development of PIRs normally are as follows:

- Determine the source(s) of the mission.
- State superior's mission and intent.
- Derive elements of own mission.
- Identify (planning) assumptions.
- Identify objective(s).

Step 1: Sources of the Mission

One's own mission is typically a sub-set of the superior's mission. It is normally contained in the superior's directive either outlined specifically for a particular operation or more generally in guidance issued to subordinate commands. Intelligence staffs supporting situational awareness within a Strategic Command can derive mission sources from the NATO Strategic Concept, Command Strategic Guidance documents, political guidance issued from the North Atlantic Council (NAC) and military guidance issued from the Military Committee (MC).

Step 2: Superior's Mission and Intent

The initial concern during mission analysis is to study the superior's mission and intent. The latter is a concise expression of the purpose of the force's activities, the desired results, and how actions will progress towards that end. The commander's intent is not a

restatement of the superior commander's concept of operations, but is rather a description of what conditions – hostile and friendly – should result from the intended operations.

Normally the content of tactical, operational, and strategic commander's intent statements will be different, but the purpose will be the same: to provide guidance concerning the military/strategic "landscape" or the situation the commander wants to exist after the assigned military mission is accomplished. In general, the higher a command echelon is, the more likely that the commander's intent will be provided in writing or in message format.

An effective technique to evaluate the superior's intent is to practice the analytical technique of "problem restatement." How you define a problem determines how you analyse it. Through the following five step process, an analyst is able to better focus on the true issue rather than the problem that was initially presented.

Steps in Problem Restatement

1. **Paraphrase: Restate the problem using different words without losing the original meaning.** Trying to say the same thing with different words puts a slightly different spin on the meaning, which triggers new perspectives and informative insights.

2. **180 Degrees: Turn the problem on its head.** Taking the opposite view not only challenges the problem's underlying premises but also directly identifies what is causing the problem.

3. **Broaden the focus: Restate the problem in a larger context.** Look at the implications of the initial problem beyond the confines of your own organization.

4. **Redirect the focus: Boldly, consciously change the focus.** Using creative thinking, look at the problem from an entirely different perspective. (e.g. from "How do we defeat the insurgents?" to "How do we eliminate the conditions contributing to the insurgency?")

5. **Ask "why": Ask "why" of the initial problem statement. Then formulate a new problem statement based on the answer. Then ask "why" again, and again restate the problem based on the answer. Repeat this process a number of times until the essence of the "real" problem emerges.** Restating a problem several different ways is a divergent analytical technique that opens the mind to alternatives.

Source: Morgan D. Jones. The Thinker's Toolkit. New York: Three Rivers Press, 1995.

Step 3: Derive Elements of Own Mission

Any mission consists of two elements: the task(s) to be done by one's own forces and their purpose. There might be situations in which a commander has been given such broad guidance that all or part of the missions will need to be deduced. This is particularly true at a Strategic Command level. Deduction should be based on an appreciation of the general situation and an understanding of the superior's intent.

A task is the job or function assigned to a subordinate unit or command by higher authority. A mission can contain single or multiple tasks based on the complexity of the activity that is envisaged. There are two types of tasks:

1. Specified Task(s): Tasks listed in the mission received from higher headquarters are specified or stated (assigned) tasks. They are what the higher commander wants accomplished. The commander's specified tasks are normally found in the "Execution" section of the order but could also be contained elsewhere.

2. Implied Task(s): After identifying the specified tasks, the commander identifies additional major tasks necessary to accomplish the assigned mission. These additional major tasks are implied tasks that are sometimes deduced from detailed analysis of the order of the higher commander and an understanding of his intent, knowledge of the operating environment, or other factors that affect the mission. Implied tasks are subsequently included in the commander's restated mission and should be limited only to those considered essential to the accomplishment of the assigned mission.

Step 4: Identify Assumptions

An assumption is a supposition on the current situation (or a presupposition on the future course of events), either or both assumed to be true without positive proof, and necessary to enable the commander, during planning, to complete an estimate of the situation and decide the course of action. Assumptions can be made of friendly as well as hostile forces. Key characteristics of assumptions are that they are reasonable suppositions, both logical and realistic.

Assumptions are necessary in preparing a mission analysis. Within NATO, the political dimension of military operations is often a larger component of the planning process than in national capitals. This is a function of the decision-making structure within the Alliance. Assumptions concerning the degree of potential Alliance involvement or interest in an international event are crucial to effective intelligence preparation.

Step 5: Identify Objectives

Objectives are those elements that provide the link between the strategic through to the tactical levels of operations. A prerequisite for a good estimate of the situation is the identification of specific, realistic, and clearly defined objectives. The scope of NATO activities has changed considerably in the last decade. While NATO has been active in operations more during this time than ever before, considerable effort has been directed at non-traditional security activities to include the Partnership-for-Peace and Mediterranean Dialogue programmes. In many cases, NATO objectives will not be military in the traditional sense but rather political, particularly in a peace support operation. When assessing NATO objectives, it is important to understand the "soft security" concerns that affect the NATO Alliance. These include humanitarian, diplomatic, economic, and general stability concerns.

Section B: Primary Intelligence Requirement (PIR) Development

The mission analysis process produces a full appreciation of the likely extent of command interest in a particular intelligence problem. Mission analysis provides a list of likely tasks, both specified and implied, that must be completed. The provision of relevant intelligence products to support the completion of these tasks is the responsibility of the intelligence staffs. Intelligence planners must remain focused on the kinds of intelligence required to support the mission. Within a NATO context, OSINT is often the only source that can be tasked directly by the commander. The categories, types and level of detail required for intelligence analysis differ from echelon to echelon. Attempts to develop intelligence products beyond the scope necessary to support the mission will likely overburden the intelligence infrastructure with too much information, which could needlessly complicate the commander's decision-making process.

Mission analysis produces the Commander's Critical Information Requirements (CCIRs)[3]. These are those elements of information that are essential for the successful completion of his mission. A sub-set of CCIRs are his PIRs. PIRs are a focused list of command's critical intelligence needs. While this list should not be extensive, it should specifically identify the intelligence necessary to enable the commander to accomplish his mission. PIRs should be ranked in order of priority.

Using PIRs as the basis, the intelligence staff is able then to develop the command's information requirements – those items of information that must be collected and processed to develop the intelligence required by the commander. If the information does not already exist, then the intelligence staff must either issue a Request for Information (RFI) to a relevant intelligence provider or initiate the development of a collection plan.

---

[3] Commander's Critical Information Requirements (CCIRs) = Friendly Force + Enemy Force + Operating Environment. OSINT can support all three of these elements and the practical exploitation of open sources should not be limited solely to the Intelligence Staff.

Direction Development Worksheet[4]

1. **What is the Mission?**
   - Who
   - What
   - Where
   - When
   - Why

2. **What is the source of that Mission?**

3. **What is the commander's intent?**
   - Purposes of the force's activities

   - Desired results

   - How actions will progress toward that end

4. **What tasks must the command do for the mission to be accomplished?**
   - Specified tasks

   - Implied tasks

5. **What are the objectives?**
   - Political
   - Military
   - Economic
   - Social
   - Religious
   - Racial
   - Psychological
   - Environmental

6. **What are the Primary Intelligence Requirements (PIRs)?**

---

[4] This Direction Development Worksheet can be used to aid in the preparation of PIRs prior to the initiation of the collection planning process.

CHAPTER III: COLLECTION

Section A: Collection Planning

Collection Management

The Collection Management process consists of Information Requirements Management (IRM), Collection Planning, and Collection Coordination and Execution. The complexity of each is dependent on the capabilities and size of the organization. Collection Management usually begins with determining and sorting information requirements (discussed in the previous chapter). With a thorough knowledge of the capabilities of the available collection assets, the Collection Manager decides on the best means to satisfy the information requirements. He then constructs a Collection Plan and directs the execution of those collection activities. The Collection Manager will repeat the process if the requirement is still outstanding or if the collection was unsuccessful.

The Collection Manager in a multi-disciplined intelligence organization would determine what type of collection assets to place at what locations during what periods of time in order to have the greatest chance of acquiring the desired information (e.g., image, electromagnetic emission, acoustic event, verbal utterance, or text document).

Making a Collection Plan

Collection Planning lays out the most effective "what, where, and when" collection options in a way that matches each information requirement to a potential collection solution:

Information Requirement = What to Collect + What Collector + Where to Collect + When to Collect

Often, more than one collection solution is assigned to a single requirement in order to improve the odds of collecting quality data. Collection Plans are typically a matrix or spreadsheet with the three key pieces of information arranged in some graphic format. For example, different reconnaissance aircraft could be on the vertical axis, the time of flight on the horizontal axis, and the desired target of collection at the intersection.

An Internet Collection Plan

If the analyst's browser connected to the Internet is the only collection asset being considered, the collection planning process is still useful. The plan will save time and improve efficiency by keeping the collector focused on exactly what he is looking for.

- **Internet Location**
  A collection plan would match each Information Requirement with a location and time to collect from the Internet. In this case, the location is the server's URL or IP address and perhaps the details about how to access the site (i.e., password, drill downs, use site's search engine, etc.). To improve the chances of finding quality information or multiple pieces of supporting data, multiple sites may have to be collected against. There will be cases when a reliable site is already known and other times when the starting location will be a preferred search engine fine-tuned with key words and other specifications.

- **Time to Collect**
  The time dimension for collection may not be as obvious as location. Many news servers and indexing servers only keep information posted for a set amount of time before it is replaced by fresh information[5]. Some move the old information off to archives and others just delete it. For this reason, the Internet Collection Plan must include when and how often to look at the desired sites.

An Internet Collection Plan may look like this example:

Information Requirement	Site URL or IP	Access Details or Key Words	Search Time or Frequency
Blue Navy Exercise	http://www.janes.com/	Password & Naval Review	Archived by month
	https://portal.rccb.osis.gov/	Password & Exercise/Navy	Daily updates
Terrorist Bombing	http://www.msnbc.com/	Terror	Within 48 hrs
	www.alltheweb		
	http://www.ict.org.il	Background on terror organizations	Frequent updates

---

[5] For example, www.alltheweb.com indexes 3,000 news sources into a searchable database. Regrettably, it keeps this archive only for 48 hours.

The Internet Collection Planning Steps

The basic four steps to construct an Internet Collection Plan are as follows:

       Step 1: Determine Searchable Information Requirements
       Step 2: Determine Best Sites or Search Strategy
       Step 3: Identify the Details to Access or Find Specific Information
       Step 4: Determine Search Time Constraints

Step 1: Determine Searchable Information Requirements

Often, the Information Requirement as stated by the Commander (or other intelligence consumer) may not be in a form suitable for Internet searching. Sometimes the wording is too broad, too narrow, too vague, or consists of uncommon jargon. PIRs and other Information requirements are usually very broad and may need to be broken down into smaller, more specific information requirements for the Collection Plan. These are the Essential Elements of Information (EEIs) although there are many names and terms associated with this concept. A classic example is the Commander's PIR "indications of hostile intent." A search for hostile and/or intent will probably result in nothing found. That PIR may be turned into a series of possible component Information Requirements such as: Indications that Blueland is preparing for war, on heightened alert, or has improved their air defense readiness.

Step 2: Determine Best Site or Search

In many cases there will be one or more known reliable news or specialized web sites that you frequently refer to for information similar to what you are seeking. A good example is Janes International Defense Review for information about new military hardware and its use. You may decide to go directly to that site or use a service that searches a variety of journals and periodicals such as FACTIVA or LEXIS-NEXIS. For current events you probably have your favorite news sources such as BBC, CNN, or Mid East Newsline. Another option is to begin to search the web with a dependable search engine like AlltheWeb.com.

Step 3: Identify the Details to Access or Find Specific Information

As the step title states, the details needed to find the specific information may be of two different types: access and things that help you narrow down the location.

You may be going to a known site for which you are a registered user or subscriber and a login ID and password are required to gain access. Sometimes you can get away with using the same password and ID but often they are assigned randomly.

When you go to a known site or search engine, there are usually ways to get closer to the information you actually need. This could be by going to a sub-page, selecting a category, entering key words or search dates, etc. These will be covered in greater detail in the next section on search methods. To use an example of searchable keywords, let's go back to the hostile intent PIR from step 1. From such a vague PIR, a list of adversarial nations, their military leaders names, and words like heightened alert, improved readiness, forward deployed, and war preparations could be searched on to find resulting information that would reveal hostile intent.

Step 4: Determine Search Time Constraints

As mentioned above, current events information is perishable on web sites and many indexing servers. Some move the old information off to archives and others just delete it. It must be determined whether the sites need to be collected on several times each day, at least once per week, or whenever you get around to it depending on the priority of the requirement and refresh rate of the site. Newspapers and journals usually have archives but some require payment or a different means of access.

Using the Internet Collection Plan

Now that you have constructed a Collection Plan, it can be used by you or someone else to methodically locate and satisfy each of your Information Requirements or to identify those requirements that must be satisfied by another means.

Standing Requirements

Time spent in preparation will significantly increase the productivity of a research staff. Standing requirements such as INTSUM preparation or recurring reports can be templated so that the collection effort can be treated almost as a mechanical process. With an established collection plan supported by a report format, it is possible to quickly train staff to collect and prepare open source reporting in a standardized form.

Maintaining a dynamic collection of Internet links to support research, while time-consuming, is an efficient management tool. Link tables with supporting search terms and revisit advice can be built for all recurring tasks within an organization.

Internet Collection Plan

Priority Intelligence Requirement and Information Requirements	Key Terms	Sources to be Employed	Time to Collect	Remarks
1.				
2.				
3.				
4.				

Section B: Search Strategies

Search Methodology

Prepare Before You Search

The world of online resources to support your research is constantly expanding. Thus before delving into a search headfirst, it is best to consider your options. Online resources include:

- **Library catalogs**
  Most libraries have catalogs of their holdings. The majority of state and educational institutions' libraries are members of library consortiums, which allow you to search the member institutions' collections through their search engine. You can use this function to locate more difficult to find resources such as rare books, journals, magazines, etc.

- **Reference databases**
  These include encyclopaedias, periodical indexes, and full-text resources, which the library licenses from publishers for your use. Use these databases to identify (and read) articles on a variety of topics.

- **Internet resources**
  These are Internet search tools that are available to everyone. It is important to note, however, that a growing number of high-quality databases can be accessed via the Internet – but with a cost associated with their usage.

Because these resources are created by different organizations, they don't all look or function in the same way. The good news is that there are key searching skills that can be used in *all* of these online resources to help you connect quickly to the information you need. This section describes **six steps for successful searching.**

Step 1: Identify Key Concepts
Step 2: Identify Possible Search Terms
Step 3: Decide Which Method to Use To Search
Step 4: Construct Your Search
Step 5. Limit Your Search
Step 6. Refine Your Search
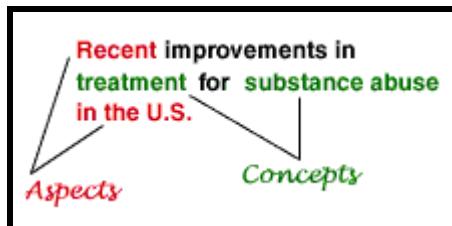
Step 1: Identify Key Concepts

The key to successful searching is good preparation. There are two parts to this first step. First, before starting to search, take just a moment to analyse your search topic. Write a sentence describing what you are looking for. Then, second, identify:

1. The main concepts (Hint: these are usually nouns rather than verbs, adverbs or adjectives)

In the illustration below, the key concepts are "substance abuse" and "treatment."

2. Any aspects of the topic (such as time, place) that will be useful for focusing your research.

In the illustration below some aspects are "recent" and "U.S."

Recent improvements in treatment for substance abuse in the U.S.

Aspects   Concepts

Step 2: Identify Possible Search Terms

Next, make a list of words you can use to express each of your main concepts. As you do this, consider:

1. Are there narrower or broader terms you want to include in your search?

For example, "substance abuse" is a broad concept. It can involve misuse of various kinds of substances, such as alcohol, drugs or tobacco. Each of these categories also has narrower aspects, as shown in the table below.

Broad term:	substance abuse		
Narrower:	alcohol	drugs	tobacco
Narrowest:	beer whiskey wine	cocaine heroin marijuana	cigarettes cigars snuff

2. What about **synonyms**? Are there different ways that your concept can be described? For example, some synonyms for "drug" are "dope, narcotic, opiate."

Often you may be able to think of synonyms, but you can also use a thesaurus to locate synonyms. There are even online sources like the *Wordsmyth Educational Dictionary - Thesaurus* (http://www.wordsmyth.net/), which can be of great assistance.

Topic Development Worksheet

The following worksheet acts as a useful tool to develop a specific topic from general concepts to specific search terms. Time spent thinking about a search before any keys are pushed can significantly increase search accuracy and completeness.

Topic Development Worksheet

1. Name of topic, and what do you want to know about the topic:

2. Spell out the topic: (Words, acronyms, abbreviations)

Generic, simple terms	Obscure, specific terms

3. Make a list of "who" might publish such information
(Industry associations, government agency, NGO's, user group etc)

Step 3: Decide Which Method to Use To Search

The two most common methods of searching online sources are by subject and keyword. However there are also phrase, concept, and natural language searches.

1. **By SUBJECT**

    How: Using standard terms or "subject headings" that have been identified by an editor to represent the *main focus* of a document.

    Where: Library catalogs and most reference databases can be searched by subject. Indexes of web sites (like *Google* or *AltaVista*), however, do not have this feature.

    Efficiency: Subject searching is a precise (and thus efficient) method for finding information. It will help you find relevant information regardless of the varying terminology that different authors may use to describe a topic.

    Best Use: Because you can search with precision (and not retrieve unrelated information), use this method when your research topic is broad (such as "substance abuse") or ambiguous (such as "Columbus" -- do you need information about a city in Ohio or Christopher Columbus?)

Requirements: You must translate your search concepts into the subject vocabulary used by the database. Sometimes it is difficult to identify the correct subject terms. Also, the terms that work in one database may not be used by another (as shown below), so you should check the subject list ("thesaurus") for the database you are using.

Database:	Thesaurus:	Subject Heading:
OSCAR	Library of Congress Subject Headings	Substance abuse
Medline	Medical Subject Headings	Substance-related disorders
PsycINFO	Thesaurus of Psychological Index Terms	Drug abuse

2. **By KEYWORD**

How: Using words that may occur *somewhere* in a document, such as the title, description (abstract) or full-text of the resource itself.

Where: This method is used when searching web indexes. Library catalogs and reference databases also allow keyword searching.

Efficiency: Keyword searching is a less precise (and often less efficient) method for finding information. Because a keyword search looks at all of the words contained in a document (not just subject headings), it will find more results for you to sift through, and many may not be relevant.

Keyword searching also won't distinguish between different meanings. For example, if you want information about Turkey (the country), you will probably also find resources about turkey (the bird) mixed together in the results of a keyword search.

Best Use: Use keyword searching when your research topic is specific (such as "substance abuse during pregnancy") or not much has been written on it.

If a database allows both keyword and subject searching, you can also use keyword searching to identify subject headings when you are not sure which terms are used in that database to describe your topic. Skim results returned by a keyword search and find an item that looks useful. Then use a subject heading from that item in a new search by subject, in order to find more information on your research topic.

Requirements: You must construct a search statement if using more than one keyword. Your statement will use "operators" to connect search words. Various operators produce different effects.

EXAMPLE: Keyword Search in Library Database

The illustration shows a record found by a search in *Periodical Abstracts* for the keywords: substance abuse pregnancy

- "substance abuse" was found in both the journal title and the subject heading.

- "pregnancy" was found in another subject heading.

Author	Farrow, James A
Title	Pregnant adolescents in chemical dependency treatment: Description and outcomes
Appears In	*Journal of Substance Abuse Treatment*. v16n2 Mar 1999. p.157-161
Abstract	This study examines the treatment, maternal and infant outcomes of pregnant adolescents (16-19 years) enrolled in an adult perinatal chemical dependency treatment program. Twenty-one adolescent subjects were compared to 323 adult women (mean age, 27.4 years) after entrollment into a randomized treatment trial consisting of intensive outpatient or short-term residential conditions.
Subjects	Teenage **pregnancy** **Substance abuse** treatment

EXAMPLE: Keyword Search in Web Index

The illustration shows the first result found by a search in *FAST* for the keywords: substance abuse pregnancy

- "substance abuse" and "pregnancy" were both found in the web page title.

- these words were also found in the page text (they are highlighted in a lighter colour).

Since this is a web index, there are no subject headings that can be used to extend your search.

All the Web, All the Time™   fast

82326 documents found - 0.361 seconds search time

1 Pregnancy Riskline of Utah*--Substance Abuse
It is difficult to obtain an accurate number of women who use substances of abuse during pregnancy. However, it is important to remember that substance abuse is an illness that can affect individuals from all walks of life. Not only do women not report the use of illicit substances during pregnancy for fear of prosecution, but also the women who are targeted for drug testing tend to be a selected population, poor and non-white. It is of interest that around 1991 there were approximately 4 million substance abusing women in the United States. Of that, 250,000 were pregnant with a total of 25
http://www.pregnancyriskline.org/cfhs/he/prl/abuse.html

Step 4: Construct Your Search

At this stage, you have identified main search concepts, developed a list of search words related to each concept, and chosen a method (subject or keyword).

With the keyword method, you can type a few search words and get results, but the outcome of this approach is unpredictable and often unsatisfactory. To perform an effective keyword search, you must formulate a search statement. This involves:

a. Selecting operators to connect your search terms
b. Identifying any search terms that are phrases
c. Deciding whether to include word variants
d. Organizing any complex search statements using parentheses.

a. Select Operators to Connect Search Words

There are several different types of "operators" that can be used to connect search words. Sometimes you will type these operators along with your search words into a search box. Sometimes you will select an option for connecting your search words from a pull-down list, like that shown in the illustration below.

Use Boolean Operators

Some databases use special connector words called Boolean operators to combine search terms. The most frequently used Boolean operators are: **AND, OR, NOT**.

Search for:	What happens?
addiction **AND** treatment	**requires** that **ALL** of these words be present in results; use AND to connect concepts
cocaine **OR** crack	**ANY** of these words can be present in results; use OR to connect synonyms
drugs **NOT** prescription	**excludes** words from results

Use Mathematical Operators

Some databases allow you to use either Boolean operators (words) or mathematical operators (symbols) to combine search terms. These mathematical operators are the **plus** symbol (**+**) and the **minus** symbol (**-**).

Search for:	What happens?
+addiction +treatment	**requires** that **ALL** of these words be present in results, like the Boolean operator AND.
+drugs -prescription	**excludes** words from results, like the Boolean operator NOT.

NOTE: Spacing counts. When constructing search statements using mathematical operators, be sure to format them correctly, so that the search works as intended:

- DON'T leave a space between the math symbol and your search word.

- DO leave a space between each element (symbol plus search term).

Be Consistent

Use either words or symbols as operators in your search statement. Don't mix them together.

> *Correct usage:* +addiction +treatment
> *Incorrect usage:* +addiction AND treatment

Implied Operators

What happens if you don't use any operators? In many databases, one of the Boolean operators (AND or OR) is implied and thus is supplied automatically by the system. *There is no standard for this, so you must check the database HELP pages to find out which operator is implied.*

b. Identify Search Terms that are Phrases

When constructing a search statement, it is important to identify any words that should be considered a phrase. Online sources use various methods for phrase searching. Depending on the database, you may be required to:
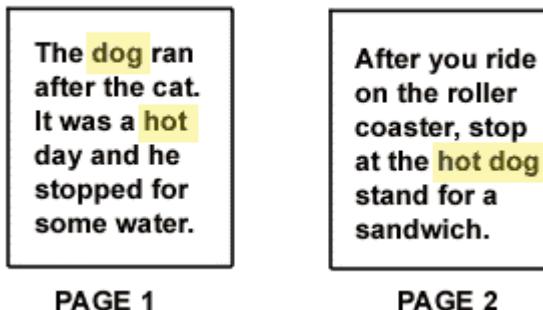
1. Use **quotation marks** to enclose an **exact phrase**
2. Choose the phrase search option from a **drop down menu**
3. Use a **Boolean operator** such as **NEAR** or **ADJ** to indicate **proximity** of terms

Search for:	What happens?
substance **NEAR** abuse	finds both words near each other (often in the same sentence), in **any order**. This search will find the exact phrase "substance abuse" as well as the phrase "abuse of a controlled substance."
substance **ADJ** abuse	finds both words next to (adjacent to) each other, in **any order**.
"substance abuse**"**	both words together in this **exact order**.

NEAR and ADJ are useful Boolean operators, but they are not universally supported by all databases. If they are available in the database you are using (check HELP to find out), use NEAR or ADJ instead of quotation marks when you are not sure of the order of the words in a phrase.

Note that phrase or proximity searching is more restrictive than using the AND operator. Using quotation marks around an exact phrase or the ADJ operator between

words that form a phrase will produce smaller, more precise results, but not always more accurate. Without care, use of these search techniques may narrow your search excessively.

PAGE 1   PAGE 2

Example:

- Search for: **hot AND dog** will find both pages 1 and 2 in the illustration above.

- Search for: **"hot dog"** or **hot ADJ dog** will find only page 2.

Summary

The table below summarizes the effect that various operators will have on your search results, from least restrictive (producing the biggest set of results) to most restrictive (producing the smallest set of results).

Operator:	*Example:*	*Web Search Results*
OR	peanut OR butter OR cookies	1,499,578 pages
AND (or plus sign)	peanut AND butter AND cookies	33,534 pages
NEAR	peanut NEAR butter NEAR cookies	10,591 pages
ADJ	peanut ADJ butter ADJ cookies	5,291 pages
phrase indicator	"peanut butter cookies"	3,838 pages

To find recipes for peanut butter cookies, for example, the most efficient search statement is:

Using Boolean operators: "peanut butter cookies" AND recipes
Using mathematical operators: +"peanut butter cookies" +recipes

c. Include Variant Forms of Search Words

Many databases execute searches quite literally. If you search for the singular form of a word (such as cat), the plural form (cats) will not be found. But often databases have a **wildcard** feature that you can use to find variant forms of your search words.

Generally, you must include some symbol, such as the question mark (**?**) or the asterisk (*****) at the end of word roots to find plural and related forms. Check database HELP pages to find out which symbol is used.

Example:
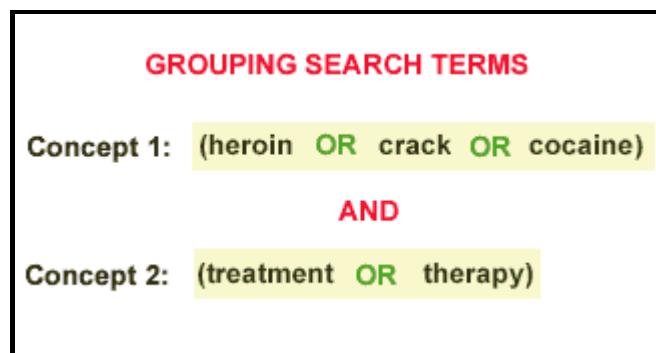
- Search for: substance**?** NEAR abuse**?**
- Will find: substance abuse, substance abuse<u>r</u>, abuse of controlled substance<u>s</u>

Some databases have a **stemming** feature that solves the problem of variant word forms. They will automatically search for "cat" if you enter the keyword "cats" and vice versa.

d. Use Parentheses to Group Search Words

It is also possible to create more complex search statements by using parentheses to group search words that should be treated in the same way. This is sometimes called "nesting" and is a handy way to incorporate **synonyms or related terms** into your search.

As the illustration below shows, you may link synonyms or related terms for each of your search concepts into a "cluster" using the OR operator. Enclose a cluster within parentheses. You can then combine clusters using the Boolean operator AND.

GROUPING SEARCH TERMS

Concept 1: (heroin OR crack OR cocaine)

AND

Concept 2: (treatment OR therapy)

At least some of the words from *each* of the concept clusters should be represented in the results for this search.

e. Recommended Strategies for Keyword Searching

- **Perform your search in stages**. First, search for the most important concepts or the most unique words (those least likely to occur in search results). Then look over your results and decide whether you need to change anything. Don't begin with search statements that are complex and elaborate. Rather, build from a simple beginning.

- **Keep phrases short**. Longer phrases are less likely to be found. For example, search for "substance abuse" AND treatment, not "substance abuse treatment programs."

- **As you review results, watch for new or alternate terms**. Incorporate them into your next search. For example, use "chemical dependency" as well as "substance abuse." Connect these search terms with the OR operator to expand your results.

Step 5. Limit Your Search

A search often returns much more information than you can possibly use. Previously, we saw that by choosing the most appropriate operator, you can produce better results. Most online databases also allow you to limit (screen) your search results to improve them. But the specific criteria that you may use to limit your search will vary.

Library catalogs and many reference databases offer different criteria or variables that can be used to limit your search results. Typically, you can limit by:

- Language
- Media or material type
- Time frame or publication date

Pre-Limit or Post-Limit? Some library catalogs and databases allow you to set up limits on the original search form (pre-limit) as well as after the initial search has returned results (post-limit). To post-limit, look for a "Limit" link or button on the search results page, like the one in the illustration below.

You searched for GUERNICA AND PICASSO. 103 records match your search - these are records 1 - 8. Limit Search Results.

Usually, **post-limiting works best**. First, try your search without any limits to see what results are produced. Then, if there are too many to review quickly (more than 2 pages), post-limit the search results by some variable that seems most relevant.

Examples:

- If your results are in many different languages, try limiting to English language only.

- If your results should be recent, try limiting by publication date.

Limiting in a Web Index

When searching a web index, some of the same limiting criteria (such as language) may be available. However, in a web Index you may also limit your search to various elements present in web pages, such as:
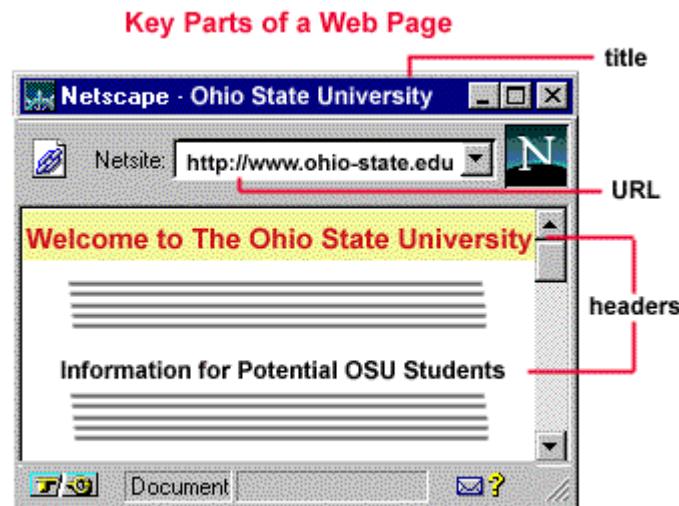
- **Page titles**

- **Headers**

- **URLs**

If your search terms are found in these key parts of a web page, it is more likely to be relevant to your needs.

Example: Web Index
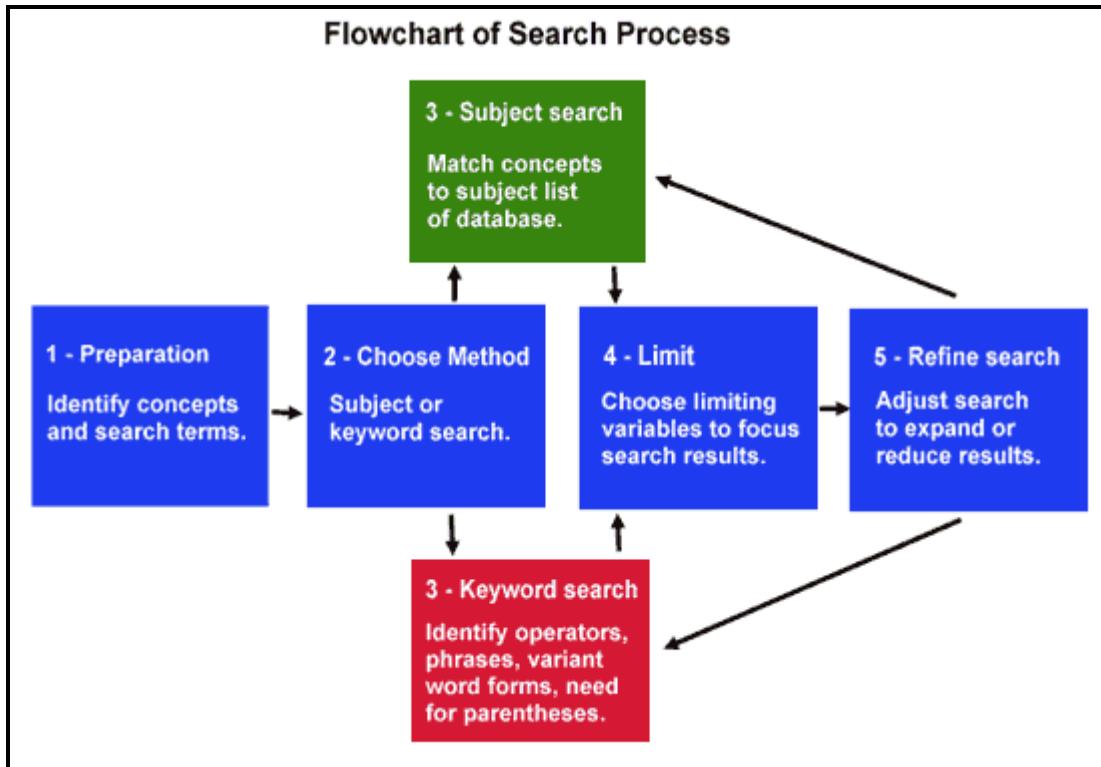
The example on the right shows a search in Google, a web index, for the phrase "Ohio State Buckeyes." By limiting this search to page titles, we will find sites that have this topic as their main focus.

The net.TUTOR tutorial on "Using Web Search Tools" provides more information on how to use the limiting feature in web indexes.

Step 6. Refine Your Search

Searching can be unpredictable. Some searches return too much information. Sometimes a well thought out search just doesn't find enough information. In any event, most online searching is "iterative," requiring that you continually refine or tune your search, as shown in the illustration below.

Flowchart of Search Process

3 - Subject search

Match concepts to subject list of database.

1 - Preparation

Identify concepts and search terms.

2 - Choose Method

Subject or keyword search.

4 - Limit

Choose limiting variables to focus search results.

5 - Refine search

Adjust search to expand or reduce results.

3 - Keyword search

Identify operators, phrases, variant word forms, need for parentheses.

Here are some proven strategies for adjusting keyword searches that find too much or too little.

Adjust:	To Narrow Results:	To Expand Results:
SEARCH CONCEPTS	**Add** search concepts, using the AND operator.	**Remove** some search concepts. Find information on the most important concept first.
SEARCH WORDS	**Remove** some search words, especially vague, ambiguous or abstract words.	**Add** more search words, using the OR operator.
SEARCH FOCUS	**Limit** search focus to a specific field or aspect (date, language, media, etc.)	**Broaden** the search focus from subject to keyword, or from one field to all fields of database.

Example:

- **Initial search:**
  "substance abuse" AND treatment

- **Too many results? Narrow by adding new concept:**
  "substance abuse" AND treatment AND Ohio

- **Too few results? Expand by adding more search words:**
  ("drug abuse" OR alcohol OR alcoholism OR "substance abuse") AND (treatment OR therapy)

Section C: Search Worksheet

1. Spell it Out

Spell-out the subject you are searching for. Write down any buzzwords, acronyms and abbreviations. Be sure the think about the "who" associated with your subject. Who would be likely to produce or publish the information you are seeking? Is there a well-known expert, university or association that specializes in your subject? The words you write down become the basis for keyword searches later on. For example; a search for anti-ship missile information might involve the following terms; missile, cruise-missile, sea-based weapons, Harpoon, Silkworm, USNI ("United States Naval Institute"), etc.

_____
_____
_____
_____
_____
_____
_____

2. Strategize - Choose your approach, which online resources, tools

Consciously decide what are the best online tools to use for each of your specific search terms. This requires that you understand that there are different strengths and weaknesses of the various online tools.

- Subject Tree (Yahoo) - Use for general terms and subject keywords only (i.e. cancer).

- Search Engines: (Alta Vista) - Use for detailed, obscure keywords. (i.e. Coetaneous T-Cell Lymphoma/Mycosis Fungicides)

- Virtual Libraries (Joe's Ultimate guide to XYZ,) - Seek out virtual libraries, if you need an in-depth guide for your subject. (i.e. OncoLink)

- Online Communities (Dejanews) - Use when you are looking for other people's opinions (i.e. alt.support.cancer, sci.med.diseases.cancer)

- Specialized Tools - There are many specialized tools, which can take you to another level of detail. These tools can quickly answer searches within their area of expertise. For example:

  o "What is John Doe's Phone number, address, directions to his house, and a listing of all his neighbour's names, phone numbers and addresses?" (See Anywho)

  o "Show me a list of the antique dealers nearest Anytown, US. (see Switchboard)

  o Who are the biggest polluters in my county? (See Scorecard)

3. Search - Get online, execute, stay focused, use advanced search features

Your search results depend greatly on how you phrase your keywords. First, be sure you are asking the appropriate level of detail for the specific tool you are using (general words at Yahoo, obscure words and specific phrases at Alta-Vista). You can focus your results when you take the time to read the help file and construct a more specific query.

4. Sift - Filter the results, Follow the leads

You may be presented with many search results and potential leads. Stop and read, before you click and waste time. Scroll up and down the entire page, and determine which are the most promising links based on the descriptions. Before you click on a link, hover over the link, and read the link's URL in the browser's feedback area (at the bottom of the screen). You will be surprised how many dead-end links you will avoid just by reading their URLs first. Finally, whenever you see 2 or more interesting links to pursue, go ahead and explore all of them simultaneously by opening multiple web browsers.
As you look through the search results, you should also be taking notes of new information that you will use in future iterations of your search.

5. Save

When you discover a great site, be sure to save this discovery, or you are doomed to repeat the search all over again. Methods of saving include:

- Add a bookmark, and organize your bookmarks into folders/submenus.

- Save a copy of the page to disk using "file" --> "save as".

- Copy and paste selected text directly from the web page into a word processing program. You may also want to copy and paste the URL into the word processor, so later on you will know where the page came from.

Summary - Take the time to think about your searches, otherwise you can waste a lot of time.

Section D: Search Tools

Search Engines

What are search engines?

Search engines are huge databases of web page files that have been assembled automatically by machine.

There are two types of search engines:

- *Individual*. Individual search engines compile their own searchable databases on the web.
- *Meta*. Metasearch engines do not compile databases. Instead, they search the databases of multiple sets of individual engines simultaneously.

How do search engines work?

Search engines compile their databases by employing "spiders" or "robots" ("bots") to crawl through web space from link to link, identifying and perusing pages. Sites with no links to other pages may be missed by spiders altogether. Once the spiders get to a web site, they typically index most of the words on the publicly available pages at the site. Web page owners may submit their URLs to search engines for "crawling" and eventual inclusion in their databases.

Whenever you search the web using a search engine, you're asking the engine to scan its index of sites and match your keywords and phrases with those in the texts of documents within the engine's database.

It is important to remember that when you are using a search engine, you are NOT searching the entire web, as it exists at this moment. You are actually searching a portion of the web, captured in a fixed index created at an earlier date.

Thus, how much earlier is an important aspect to consider when using a search engine. Spiders regularly return to the web pages they index to look for changes. When changes occur, the index is updated to reflect the new information. However, the process of updating can take a while, depending upon how often the spiders make their rounds and then, how promptly the information they gather is added to the index. Until a page has been both "spidered" AND "indexed," you won't be able to access the new information. Thus, the more often a spider checks for changes, the more accurate the search returns for that page will be.

The accuracy of your search is directly proportional to how many web pages the search engine indexes. The more web pages the engine indexes, the more accurate your search. Also, the accuracy of your search also depends on how often the search engine indexes web pages. The more often a search engine indexes web pages, the more accurate your search will be.

What are the pros and cons of search engines?

On the pro side, search engines are the best means devised yet for searching the web. Stranded in the middle of this global electronic library of information without either a card catalogue or any recognizable structures, how else are you going to find what you're looking for?

On the down side, the sheer number of words indexed by search engines increases the likelihood that they will return hundreds of thousands of irrelevant responses to simple search requests. Remember, they will return lengthy documents in which your keyword appears only once. Secondly, the vastness of the Internet means that no search engine can possibly actually search all of it. Technology may change this in the future, but as of now, even the most comprehensive search engine (Google) only covers a fraction of the entire web. Another problem with search engines is what is called the "invisible" or "deep" web. For details on the "invisible" web see the section "Searching the Invisible Web."

Are search engines all the same?

NO. Search engines use proprietary software programs to search their indexes for matching keywords and phrases, presenting their findings to you in some kind of relevance ranking. Although software programs may be similar, no two search engines are exactly the same in terms of size, speed and content; no two search engines use exactly the same ranking schemes, and not every search engine offers you exactly the same search options. Therefore, your search is going to be different on every engine you use. The difference may not be a lot, but it could be significant. Recent estimates put search engine overlap at approximately 60 percent and unique content at around 40 percent. Thus, it is VERY important to carefully consider how a particular search engine compiles its results when conducting Internet searches.

How do search engines rank web pages?

In ranking web pages, search engines follow a certain set of rules. These vary from one engine to another. Of course, the goal of all of them is to return the most relevant pages at the top of their lists. To do this, they look for the location and frequency of keywords and phrases in the web page document and, sometimes, in the HTML metatags. They check out the title field and scan the headers and text near the top of the document. Some of them assess popularity by the number of links that are pointing to sites; the more links, the greater the popularity, i.e., value of the page. Because they vary greatly, it is crucial to understand how a particular search engine ranks results. For example, if you were searching on a current events related topic, a search engine that ranks results by date would be extremely useful.

When do you use search engines?

Search engines are best at finding unique keywords, phrases, quotes, and information buried in the full-text of web pages. Because they index word by word, search engines are also useful in retrieving tons of documents. If you want a wide range of responses to specific queries, use a search engine.

Note: Today, the line between search engines and category search engines is blurring. Search engines no longer limit themselves to a search mechanism alone. Across the web, they are partnering with subject directories, or creating their own directories, and returning results gathered from a variety of other guides and services as well. One example is Google's (a search engine) teaming with Open Directory (http://www.dmoz.org/) to categorize search results.

A note on metasearch engines.

Use metasearch engines at your own risk for while they do have certain advantages (i.e. they can cover a large number of search engines in a single search), they have many weaknesses. These weaknesses include: most don't search all of the largest engines, most don't give you more than 10 records from each search engine, and most list paid listings first. The best way to demonstrate the weakness of metasearch engines is through the following example.

Caption 1. Search for: *"geologic resources" worcester* via search engines directly versus metasearch engines

	Done Directly	via *DogPile* (Results per site)	via *metacrawler.com* (Results per site)	via *search.com* (Results per site)
Google	39	0	0	0
AllTheWeb	40	0	0	0
HotBot	15	0	0	0
Alta Vista	9	0	0	2

Search Options Offered by Selected Search Engines

Feature	Search Engine
Boolean operators	AltaVista Advanced Search (http://www.altavista.com/) C4 (http://www.c4.com/) Dogpile (http://www.dogpile.com/) HotBot (http://www.hotbot.com/) Ixquick Metasearch (http://www.ixquick.com/) ProFusion (http://www.profusion.com/) WebCrawler (http://www.webcrawler.com/)
Full Boolean logic with parentheses, e.g., *behaviour and (cats or felines)*	AltaVista Advanced Search (http://www.altavista.com/) C4 (http://www.c4.com/) HotBot (http://www.hotbot.com/) Ixquick Metasearch (http://www.ixquick.com/) MSN Search Advanced Search (http://search.msn.com/advanced.asp)
Implied Boolean +/-	Most search engines offer this option
Boolean logic by template terminology	AllTheWeb Advanced Search (http://www.alltheweb.com/advanced) AOL.COM Search Options (http://search.aol.com/refine.adp) HotBot (http://www.hotbot.com/) Lycos Pro (http://lycospro.lycos.com/) MSN Search Advanced Search (http://search.msn.com/advanced.asp) ProFusion Advanced (http://www.profusion.com/) Snoopa (http://www.snoopa.com/)
Proximity operators	AltaVista Advanced Search (http://www.altavista.com/) Google (http://www.google.com/) Ixquick Metasearch (http://www.ixquick.com/)

Search Engine Guide: The following charts provide a snapshot of the features provided by various search engines as of May 2002. Familiarity with the features and strengths/weaknesses of each can radically improve the accuracy of the search results returned.

Major Search Engine – Features Guide - 2002

	All The Web	All the Web Advanced	Alta Vista Advanced	Alta vista Advanced
URL	http://www.alltheweb.com/	http://www.alltheweb.com/advanced	http://www.altavista.com/	http://www.altavista.com/sites/search/webadv
SIZE (pages)	600 million	600 million	500 million	600 million
SIMPLE BOOLEAN	*term* *-term* *defaults to an AND*	*(menu)* *+ term* *-term*	*Term* *-term* *Defaults to an OR*	Defaults to phrase
FULL BOOLEAN_	(term] term2) equals an OR			OR AND AND NOT ( )
PHRASE	(menu) " "	(menu) " "	" " some automatic	" " automatic
PROXIMITY				NEAR (= within 10 Words)
TRUNCATION			*term* (asterisk - internal or at the end of term)*	*term* (asterisk - internal or at the end of term)*
TITLE FIELD		(menu)	title: term	title: term
DATE FIELD		menu		(date range boxes)
URL FIELD		(menu)	url: term domain: term host:term	url: term domain: term host:term
"LINKS TO" A URL		(menu)	link:term	link:term
LANGUAGE		(menu)	(menu)	(menu)
MEDIA SEARCHING	Links to News, Pictures, Videos, MP3, FTP Searches	News, Pictures, Videos, MP3, FTP all have their own advanced mode	Links to Media searches image : term	image : term
CASE SENSITIVE			yes	yes
SEARCH ALL COMMON WORDS	yes	yes	yes	yes
WEB DIRECTORY ATTACHED			yes LookSmart	
CLUSTERED RESULTS	no	no	yes	yes
OUTPUT OPTIONS	standard	standard 10, 25, 50, 75, 100	Standard Customizable	standard one result per website 10, 20, 30, 40, 50 results customizable
PAID SITES FIRST	no	no	yes	yes
SIMILAR PAGES			yes	yes
ALSO SHOWN ON RESULT PAGES			"Others searched for Headlines"	"Others searched for Headlines"
OUTSTANDING SPECIAL FEATURES	Largest multimedia search Adult content filter Fast, extensive news search FTP search	Adult content filter	Images/ audio/ video search Translations Adult content filter Hit terms Highlighted	Images/ audio/ video search Translations Adult content filter

	Excite Home	Google	Google Advanced
URL	http://www.excite.com/	http://www.google.com/	http://www.google.com/advanced_search ?hl=en
SIZE (pages)	330 million	3 billion, not all fully-indexed	3 billion, not all fully-indexed
SIMPLE BOOLEAN	*Term* *-term* *defaults to an OR*	*-term* *defaults to an OR*	*(text boxes)* *defaults to an AND*
FULL BOOLEAN	AND OR NOT [?] ( )	OR	
PHRASE	" "	" "	" " (text boxes)
PROXIMITY			
TRUNCATION	automatic		
TITLE FIELD			
DATE FIELD			
URL FIELD			
"LINKS TO" A URL		link:term	link:term
LANGUAGE			(menu)
MEDIA SEARCHING	"Photos" radio button		Image Search Box
CASE SENSITIVE			
SEARCH ALL COMMON WORDS		No, but can force it with a+	No, but can force it with a+
WEB DIRECTORY ATTACHED	yes Look Smart	yes link to open directory	
CLUSTERED RESULTS	no	yes	yes
OUTPUT OPTIONS	Titles Titles & Summaries Grouped by URL	10, 20, 30, 50, 100 Results	10, 20, 30, 50, 100 Results
PAID SITES FIRST	no	no	No, but can force it with a+
SIMILAR PAGES		yes	yes
ALSO SHOWN ON RESULT PAGES	Directory hits Related searches ("Zoom in") stock and company information links Travel links etc.	Open Directory categories and sites Link to cached page Translation option Stock Quote option Link to definitions, Maps Headlines Address & phone #s	Open Directory categories and sites Link to cached page Translation option Stock Quote option Link to definitions, Maps Headlines Address & phone #s
OUTSTANDING SPECIAL FEATURES	Concept Searching News search Hit terms highlighted	Ranking based on "link" popularity Newsgroups & images Cached pages Adult content filter Covers PDF & other formats	Ranking based on "link" popularity Newsgroups &images Cached pages Adult content filter Covers PDF & other formats

	Hot Bot	Hot Bot Advanced	Lycos	Lycos Advanced
URL	http://www.hotbot.com/		http://www.lycos.com	http://search.lycos.com/adv.asp
SIZE (pages)	230 million	230 million	Uses the Fast Search database - 600 million	Uses the Fast Search database - 600 million
SIMPLE BOOLEAN	*(menu)* *term* *-term* *defaults to an AND*	*(menu)* *term* *-term* *defaults to an AND*	*Term* *-term* *defaults to an AND*	*(menu)* *term* *-term* *defaults to an AND*
FULL BOOLEAN_	OR AND NOT ( ) (Must also use menu)	OR AND NOT ( ) (Must also use menu)	(term 1 term 2) equals an OR	(term 1 term 2) equals an OR
PHRASE	(menu) " "	(menu) " "	" "	(menu) " "
PROXIMITY				
TRUNCATION		"Stemming" option		
TITLE FIELD	(menu) title: term	(menu) title: term		(box under "Link Referrals"tab)
DATE FIELD	(menu)	(menus)		
URL FIELD	(menu) domain: term	(menu) domain: term (Region menu)		(boxes, under "Page Field" tab)
"LINKS TO" A URL	(menu)	(menu)		(boxes, under "Link Referrals" tab)
LANGUAGE	(menu)	(menu)		(radio buttons under "Language" tab)
MEDIA SEARCHING	(checkboxes)	(checkboxes)	Links to "Multimedia" search	(radio button)
CASE SENSITIVE				
SEARCH ALL COMMON WORDS			yes	yes
WEB DIRECTORY ATTACHED	yes Open Directory		yes Open Directory	(link to Open Directory)
CLUSTERED RESULTS	yes	yes	no	no
OUTPUT OPTIONS	Full, brief, URL's only 10, 25, 50, 100, results from this site only	Full, brief, URL's only 10, 25, 50, 100, results from this site only	standard	standard
PAID SITES FIRST	yes	yes	yes	no
SIMILAR PAGES				
ALSO SHOWN ON RESULT PAGES	Related searches Direct Hit popularity results (top 10) Directory categories Travel links Stock links	Related searches Direct Hit popularity results Directory categories Travel links, Stock links	Reviewed sites Directory hits Links to quotes, news, travel links, etc. Matching news and shopping items Related searches	
OUTSTANDING SPECIAL FEATURES	Direct Hit popularity results	Direct Hit popularity results Search by region	Hit terms Highlighted Adult content filter	Adult content filter Downloads,MP3, news, newsgroups, etc. Hit terms Highlighted

	Teoma	WiseNut	Yahoo!
	http://www.teoma.com/	http://www.wisenut.com/	http://www.yahoo.com/
SIZE (pages)	200 million?	1.5 billion	1-2 million in directory (est.), 6m from Google
SIMPLE BOOLEAN	*term -term* *defaults to AND*	*term -term* *defaults to AND*	*+term -term* *defaults to AND*
FULL BOOLEAN_			
PHRASE	" "	" "	" "
PROXIMITY			
TRUNCATION			Automatic term* (asterisk)
TITLE FIELD			term
DATE FIELD			(menus under "Advanced Search"))
URL FIELD			u:term
"LINKS TO" A URL			
LANGUAGE		(menu on Set Preferences page)	
MEDIA SEARCHING			
CASE SENSITIVE			
SEARCH ALL COMMON WORDS	yes		
WEB DIRECTORY ATTACHED			Primarily is a directory
CLUSTERED RESULTS		yes	no
OUTPUT OPTIONS		standard	standard (Advanced search allows choose of 10, 20, 50, 100)
PAID SITES FIRST			no
SIMILAR PAGES			
ALSO SHOWN ON RESULT PAGES	Categories Expert's Links		Categories Directory hits Google hits, news, web events Links to stock quotes, news, etc. Related searches
OUTSTANDING SPECIAL FEATURES	Expert's Links	Automatic categorization of results	Hit terms Highlighted Automatic transfer of "no hits" search to partial Google

Source: Ran Hock, Online Strategies, www.onstrat.com

Search Tools

Copernic

Rather than laboriously employing a number of search engines individually in order to cast a wider net, there are a number of meta-search engines that will simultaneously search a number of sources and return a consolidated response. One of the best known of these is Copernic 2001. It is available from www.copernic.com.

The free version of Copernic searches most of the major search engines. The full version adds a number of categories including *newsgroups* to the customized list of search engines that it employs. In addition it also allows the user to search in different languages and in specific countries. Copernic can also be used to automatically update searches and provide results ready for analysis upon arrival in the office.

Deep-Web/Invisible Web

Searching the Invisible Web

The *invisible web* is composed of web pages that can be accessed via the Internet, but are not found by search engines. These are pages that are either located too "deep" in a web site for a search engine's spider to locate, are pages that a search engine cannot index because it technically can't do so, or are pages which the search engine cannot access because they lack the proper password.

In 1994, Dr. Jill Ellsworth first coined the phrase "invisible Web" to refer to information content that was "invisible" to conventional search engines. Perhaps a more appropriate term is "The Deep-Web" as most of this content resides in searchable databases, the results from which can only be discovered by a direct query. Without the directed query, the database does not publish the result. When queried, deep Web sites post their results as dynamic Web pages in real-time. Though these dynamic pages have a unique URL address that allows them to be retrieved again later, they are not persistent.

Search engines — the primary means for finding information on the "surface" Web — obtain their listings in two ways. Authors may submit their own Web pages for listing, generally a minor contributor to total listings. Or, search engines "crawl" or "spider" documents by following one hypertext link to another. Simply stated, when indexing a given document or page, if the crawler encounters a hypertext link on that page to another document, it records that incidence and schedules that new page for later crawling. Like ripples propagating across a pond, in this manner search engine crawlers are able to extend their indexes further and further from their starting points.

Thus, to be discovered, "surface" Web pages must be static and linked to other pages. Traditional search engines cannot "see" or retrieve content in the deep Web, which by definition is dynamic content served up in real time from a database in response to a direct query. Public information on the deep Web is currently 400 to 550 times larger

than the commonly defined World Wide Web. The deep Web contains 7,500 terabytes of information, compared to 19 terabytes of information in the surface Web.

The deep Web contains nearly 550 billion individual documents compared to the 1 billion of the surface Web. More than an estimated 100,000 deep Web sites presently exist. Sixty of the largest deep Web sites collectively contain about 750 terabytes of information – sufficient by themselves to exceed the size of the surface Web by 40 times.

There are several web sites dedicated to searching the invisible web. The most comprehensive of these is Direct Search, maintained by Gary Price, a reference librarian at the George Washington University, in Washington, DC and the author of "The Invisible Web." These sites also contain more information on the invisible web itself. Note that these sites are mostly directories of sites and not individual sites. Therefore, one should be prepared to cast their search net very broadly by topic area.
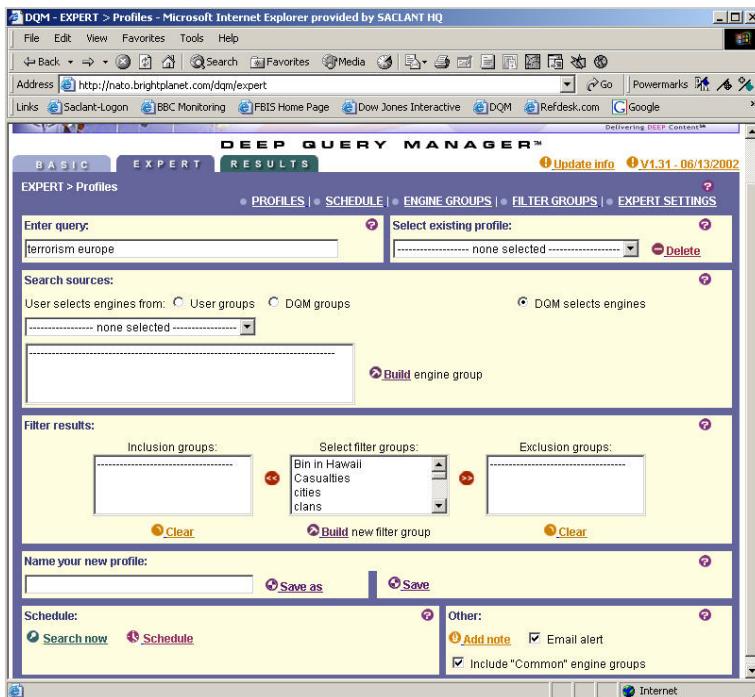
- Direct Search - http://www.freepint.com/gary/direct.htm
- Invisible Web (based on the book) – http://www.invisible-web.net/
- Invisible Web.com – http://www.invisibleweb.com/
- Complete Planet – http://www.completeplanet.com/

Deep Query Manager

In order to reach these sites in an efficient manner a specialized tool is required. Deep Query Manager is a web-based search service that enables effective access to the Deep Web. It was produced and maintained by Bright Planet of Sioux Falls, SD.

Deep Query Manager provides the means with which to search literally thousands of Deep Web sites simultaneously with the same query, providing a consolidated return of results based on document relevance. The service provides a number of other useful features including:

- The conduct of searches with a degree of anonymity through the use of Bright Planet's third-party server.

- The customizing of specific search sources based on topic parameters.

- The ability to compare results of consecutive searches on the same sources to return only the new documents.

- The capability to share search results with other analysts within your organization through the share results feature.

Deep Query Manager can run a variety of searches simultaneously. Search completion is notified via an email alert. While even a service like Deep Query Manager can take some time to search thousands of web sites, it is able to conduct many searches simultaneously thus increasing overall efficiency.

Deep Query Manager also has a scheduling function. This is useful for repetitive searches. Searches can be scheduled for completion prior to the start of the workday so that there is no waiting time prior to the commencement of analysis.

The capabilities provided in a product like *Deep Query Manager* provide two key benefits: a higher degree of confidence that a greater number of available information sources have been tapped, and; efficient use of technology to reduce repetitive tasks thus increasing the time available for analysis of the collected information.

Section E: Searching Anonymously On The Web

General

Why would anyone want to search anonymously on the Web? All the information on the Web is freely available to anyone with a PC and a connection. The answer lays not so much in people identifying who we are but, without a few precautions, we might give away our intentions and shut down what may become valuable sources of information.

There are strong arguments in favour of not hiding our presence on the Web. Some would argue that to enhance security and co-operation we should be as open as possible. This is a very strong argument but ignores the fact that some security will always be necessary. An obvious presence on the Web actually helps our own security because it means we can only apply specialised and anonymous search techniques only when and where they are really needed.

Being anonymous on the Web may not necessarily involve deception. It is quite possible to surf the Web without openly identifying your identity, purpose or intentions. This is simply a case of 'I won't tell you unless you ask'. There may be occasions when you will want to communicate with someone without using your real name. These occasions should be rare and should be assessed individually.

Don't leave yourself open to attack

Before you even start surfing anonymously make sure you don't leave your Internet connection open to attack. You may take all the precaution necessary to hide your intentions and identity whilst surfing, but if all the sites that you have visited and all the information you have downloaded is stored on your PC and available via your open connection, then you are vulnerable.

There is an argument for the use of a firewall to stop hackers at the front door, but remember that there hasn't been a firewall yet that wasn't eventually cracked. The very expensive firewalls do a good job but it is unlikely that you would want to spend so much money for a simple Internet connection. Besides if you wanted to remain anonymous on the Internet an expensive firewall is not the way to do it. It would highlight the fact that you had something to hide. The same argument applies to the cheaper firewalls. Because they are cheap they are also vulnerable. Hackers know how to get in and often see areas with firewalls as a challenge.

Possibly the best security is to remain anonymous and look just like everyone else. By doing this, if a hacker chooses to attack your Internet connection whilst you are online, they would find………nothing. The hacker would probably then get bored and never bother you again.

So how do you create an anonymous PC? There are a few simple steps and rules to follow:

- Disable anything that records your activity. If using MS Internet Explorer turn off the cookies, clean out the history folders, and routinely remove cached files.

- Use removable storage media to save any downloaded files to.

- Only use your Internet PC for surfing. Do not use the word processor for business or personal letters.

- Make sure all your connection details are anonymous.

- Ensure the set up of your system is as standard as possible.

- Do not use your ISP Email for anything other than anonymous traffic.

- These few simple steps should help to keep your system clean and anonymous to the casual hacker.

Do not leave a footprint

When you surf the Internet you cannot fail to leave a footprint. A footprint is an electronic signature that identifies you as a unique identity on the Internet during your current session. Most ISPs now issue a new signature to you each time you log on to surf. But while you are surfing during a session after log on, every site you visit retains your electronic signature. If you are trying to find information on a sensitive subject it is possible to carry out an analysis of the sites you visit and the subjects you are searching for. It would be sensible to log off and on again a number of times during a sensitive search.

Although an ISP may provide you with a new signature, part of that signature will identify the ISP. If your organisation is large and has its own ISP this will identify your organisation. It is always better to go through a civilian ISP whenever possible.

It is possible to use a number of different ISPs. These days there are a huge number of free ISPs available. Each country has its own list of free ISPs and details of these can be obtained via the Internet. It is possible to hide the country from which you are searching from by dialling up an ISP in another country and beginning your search from there. It is almost impossible for a hacker to identify which country your call originated from because Telecom companies take their security very seriously. There is one thing that may identify your country of origin and that is the date and time of your search. When you surf the date and time of your PC is stamped on the search as part of the electronic signature. If this does not match your ISP time it may indicate that you are trying to hide something, so make sure your PCs time is set to the time zone of the country of the ISP you are using.

Traffic analysis

Every web site has the capability to log the number of visitors to its site and the electronic signature of the visitor. While you may be able to hide your identity, you cannot hide the fact that you have visited the site. If the number of visitors to a significant site increases dramatically then this may be an indicator that there is new or renewed

interest in the subject of the site. Such a site may be set up deliberately to identify interest, for example an obscure terrorist related site. The way to combat this is to ensure that trained personnel, in a central location, do all sensitive searches. This will ensure that searches are done quickly and without repetition. The security education of all personnel who have access to the Internet is also a very important factor.

Contact with others

There may be occasions when you want to communicate with others to solicit information. In most cases it is beneficial to explain who you are and ask for help or information. There may be other occasions when you will not want others to know exactly who you are or who you work for. The reasons for this must be decided case by case. It is reasonably easy to create an anonymous persona on the Web but the following points should be noted.

First, it is better to employ discretion rather than deception when soliciting information on the Web. This will be less publicly embarrassing later and will make an explanation of your action more reasonable.

Second, an anonymous persona should only be used for occasional requests for information. Any development of a relationship using the Internet should be discouraged. This is the field of other specialists, typically in the realm of HUMINT. Without proper control, such practices can lead to embarrassment.

Conclusions

- It is better to be discrete when searching on the Internet rather than employ deception.

- When searching discretely you are hiding your intentions.

- Practical measures should by applied sparingly and only with good reason.

CHAPTER IV: PROCESSING

Section A: Source Evaluation

Critically Analysing Information Sources

You can begin evaluating a physical information source (a book or an article for instance) even before you have the physical item in hand. Appraise a source by first examining the bibliographic citation. The bibliographic citation is the written description of a book, journal article, essay, or some other published material that appears in a catalog or index. Bibliographic citations characteristically have three main components: author, title, and publication information. These components can help you determine the usefulness of this source for your paper. (In the same way, you can appraise a web site by examining the home page carefully.)

a. Initial Appraisal

Author

1. What are the author's credentials--institutional affiliation (where he or she works), educational background, past writings, or experience? Is the book or article written on a topic in the author's area of expertise? You can use the various *Who's Who* publications for the U.S. and other countries and for specific subjects and the biographical information located in the publication itself to help determine the author's affiliation and credentials.

2. Have you seen the author's name cited in other sources or bibliographies? Respected authors are cited frequently by other scholars. For this reason, always note those names that appear in many different sources.

3. Is the author associated with a reputable institution or organization? What are the basic values or goals of the organization or institution?

Date of Publication

1. When was the source published? This date is often located on the face of the title page below the name of the publisher. If it is not there, look for the copyright date on the reverse of the title page. On web pages, the date of the last revision is usually at the bottom of the home page, sometimes every page.

2. Is the source current or out-of-date for your topic? Topic areas of continuing and rapid development, such as the sciences, demand more current information. On the other hand, topics in the humanities often require material that was written many years ago. At the other extreme, some news sources on the web now note the hour and minute that articles are posted on their site.

Edition or Revision

Is this a first edition of this publication? Further editions indicate a source has been revised and updated to reflect changes in knowledge, include omissions, and harmonize with its intended reader's needs. Also, many printings or editions may indicate that the work has become a standard source in the area and is reliable. If you are using a web source, do the pages indicate revision dates?

Publisher

Note the publisher. If the source is published by a university press, it is likely to be scholarly. Although the fact that the publisher is reputable does not necessarily guarantee quality, it does show that the publisher may have high regard for the source being published.

Title of Journal

Is this a scholarly or a popular journal? This distinction is important because it indicates different levels of complexity in conveying ideas. Or you may wish to check your journal title in the latest edition of *Katz's Magazines for Libraries* for a brief evaluative description.

b. Content Analysis

Having made an initial appraisal, you should now examine the body of the source. Read the preface to determine the author's intentions for the book. Scan the table of contents and the index to get a broad overview of the material it covers. Note whether bibliographies are included. Read the chapters that specifically address your topic. Scanning the table of contents of a journal or magazine issue is also useful. As with books, the presence and quality of a bibliography at the end of the article may reflect the care with which the authors have prepared their work.

Intended Audience

What type of audience is the author addressing? Is the publication aimed at a specialized or a general audience? Is this source too elementary, too technical, too advanced, or just right for your needs?

Objective Reasoning

Is the information covered fact, opinion, or propaganda? It is not always easy to separate fact from opinion. Facts can usually be verified; opinions, though they may be based on factual information, evolve from the interpretation of facts. Skilled writers can make you think their interpretations are facts.

Does the information appear to be valid and well researched, or is it questionable and unsupported by evidence? Assumptions should be reasonable. Note errors or omissions.

Are the ideas and arguments advanced more or less in line with other works you have read on the same topic? The more radically an author departs from the views of others in the same field, the more carefully and critically you should scrutinize his or her ideas.

Is the author's point of view objective and impartial? Is the language free of emotion-arousing words and bias?

Coverage

Does the work update other sources, substantiate other materials you have read, or add new information? Does it extensively or marginally cover your topic? You should explore enough sources to obtain a variety of viewpoints.

Is the material primary or secondary in nature? Primary sources are the raw material of the research process. Secondary sources are based on primary sources. For example, if you were researching Konrad Adenauer's role in rebuilding West Germany after World War II, Adenauer's own writings would be one of many primary sources available on this topic. Others might include relevant government documents and contemporary German newspaper articles. Scholars use this primary material to help generate historical interpretations--a secondary source. Books, encyclopaedia articles, and scholarly journal articles about Adenauer's role are considered secondary sources. In the sciences, journal articles and conference proceedings written by experimenters reporting the results of their research are primary documents. Choose both primary and secondary sources when you have the opportunity.

Writing Style

Is the publication organized logically? Are the main points clearly presented? Do you find the text easy to read, or is it stilted or choppy? Is the author's argument repetitive?

Evaluative Reviews

Locate critical reviews of books in a reviewing source, such as *Book Review Index*, *Book Review Digest,* or *Periodical Abstracts*. Is the review positive? Is the book under review considered a valuable contribution to the field? Does the reviewer mention other books that might be better? If so, locate these sources for more information on your topic.

Do the various reviewers agree on the value or attributes of the book or has it aroused controversy among the critics?

Source: http://www.library.cornell.edu/okuref/research/skill26.htm - LinkAuthor

Determining the Source of Web Pages

Step 1: Study the URL

protocol://computer.domain.name/pathname/filename.ext

The URL of every web page is displayed at the top of your web browser. Get into the habit of always reading the URL first before looking at the web page. (see How to Read a URL immediately following this section). Often times you have arrived directly at a web page via a search result or another hyperlink. You may be several levels "deep" in a particular web site. Study the URL and determine what kind of web server the page is hosted within. For example, delete the latter half of the URL and visit the main page at "computer.domain.name"

- If "company.domain.name" looks to be an Internet provider, or web hosting service, then the specific web page you were viewing probably belongs to a user of that service. At this point, your focus is strictly on the individual user. Go back to the URL and see of there is a logical breaking point in the URL. For example: **http://company.domain.name/pathname/filename.ext** can be shortened to **http://company.domain.name/pathname** If Pathname is a User_ID_NAME, then this approach should bring you to home page for a users' directory.

- If "company.domain.name" is some kind of specific organization that you want to learn more about, then proceed to the next steps.

Step 2: Do a "whois" on the domain name

All domain names on the Internet are registered with "domain name registrars" Domain name registrars are entities, which have been allocated the authority to register names for a specific subset of domain names. Most domain name registrars provide a "whois" function, where you can ask "whois domain.name" and they will tell you who has registered that domain name.

Useful Whois locations include:

- NSI Registry. Integrated database for all .com, .net, .org, .edu).

- Popular registrar's whois: Network Solutions, Register.com, and many other registrars.

- *AllWHOIS* (- Another nicely done integrated WHOIS interface to most WHOIS servers worldwide.

- Whowhizz - allows domain name look-ups of many countries.

- *Geektools* - provides a whois interface that will automatically query the right registrar - very useful when dealing with many types of international domain names.

- Whois.net - you can do a key-word search of domain names, with the results linking to whois.

- An extensive table of domain name registrars around the world is available at http://www.norid.no/domreg.html. Two letter country codes can also be found at the same site.

- US .MIL Domains can be researched at http://www.nic.mil/dodnic/

- Domainwatch -has some nice features, but seems to work for only some .com, .net, and .org names. Whois data records are linked to other search results such as: traceroute, map info, and list of all websites that link to the queried domain.

- Betterwhois - search the shared registry first and then the specific registrar.

- Network-tools.com. - integrates many functions into one page.

- Domainsurfer and Netcraft - Perform a keyword search against all domain names. This is helpful when searching for a new unique domain name.

Regional Internet registries handle the allocation of IP Numbers. Using their "whois" you can search 'based on IP Numbers, autonomous system numbers (ASNs), network-related handles, and other related Points of Contact (POCs)

- ARIN Whois - American Registry for Internet numbers.

- Ripe Whois - European IP network Coordination Center

- APNIC Whois - Asia Pacific Network Information Centre

To learn more about Domain Name structure and to locate domain name registrars for 2-letter domains (.UK, .JP, etc) explore the following links:

- Domain Name Registration (RFC 1591 - Domain Name System Structure and Delegation – on-line at http://www.isi.edu/in-notes/rfc1591.txt)

- Another list of Internet registrars from around the world (includes name, addresses and phone numbers) can be found at http://www.itu.int/net/cctlds/icc-a-z.htm.

- The US Department of Defense Network Information Centre (DoD NIC) manages the .mil generic Top Level Domain (gTDL): http://www.nic.mil/.

- The Internet Corporation for Assigned Names and Numbers (ICANN – http://www.ican.org/) will now have responsibility for the IP address space

allocation, protocol parameter assignment, domain name system management, and root server system.

- An excellent primer on Domain Name System (DNS) is can be found in Chapter 2 of DNS and BIND, 3rd Edition by O'Reilly & Associates. This book is also available on line at http://www.oreilly.com/catalog/dns3/chapter/ch02.html.

- A very comprehensive, yet layman friendly explanation of DNS is found "DNS - The Internet's Directory Service" on the Networking Next web site at http://www.networkingnext.com/basicnetworking/dns.html.

Source: http://navigators.com/whois.html

Step 3: Perform a Traceroute to the host name

Knowing who owns the domain may not satisfy your curiosity. You may also be interested in **where** is the web server located, and **how** is it connected to the Internet. There is a network utility called Traceroute, which is often used to trouble shoot network connections. In a Unix or Windows environment, Traceroute can be used to determine the specific network route taken from your workstation to reach a specific remote host. (Dos command is: "tracert sitename.com") Fortunately, there are many Unix systems on the Internet that allow us to originate a Traceroute from their location to any other location that you specify. You should recognize that a web server does NOT have to be hosted at an organization's location, but may be hosted with some Internet provider.

Step 4: Read the web-page and follow-up with the point of contact (if any)

Many webpages include point of contact information. You can also send an email to the page owner and ask, "Who are you". You should also examine the HTML Source code of the web pages - Many web-authoring programs include the author's name within meta_tags, which are included near the beginning of a web page. You may also learn about the organization by searching for that organization's name/domain name in message archives such as Google Groups (http://groups.google.com/).

Finally, if none of these approaches reveal any information about the web page/site... then you will want to "search upstream" of the web page to see who else links **towards** the web page you are interested in. (Explained in Annex A – More information on Source Evaluation)

Source: http://navigators.com/sesseval.html

How to Read a URL

URL's typically have the following format:

protocol://computer.domain.name/pathname/filename.ext

1. "**protocol://**" - This defines what Internet protocol is required to reach the online resource. Commonly used protocols include:

    - http:// - Hypertext Transfer Protocol - used to access a server that is supporting the WWW protocol (i.e. web server) - commonly used for downloading web pages and associated imbedded elements

    - ftp:// - File Transfer Protocol - used to download a file from a server supporting the FTP Protocol - commonly used to download a software program

    - news: - Used to access a Usenet newsgroup from your news server - Your web browser must be configured to access a specific news-server

    - telnet:// - Establish a telnet session (terminal emulation) to the specified host (often a VT100 Session)

    - mailto: Initiates an outgoing email message to the address specified

2. "**computer.domain.name**" - The domain name of the server where the information is located (can also be the server's IP number)

3. "**/pathname/**" - Usually consists of directory/subdirectory names. This defines where on the server's hard disk to look for the information.

4. **"filename.ext"** - The name of the desired file. If no specific filename is indicated, a file called "index.html", "default.html", or "home.html" may be downloaded if present. The ".ext" file extension cues the web browser on how to handle the downloaded file. The web browser can display some file-types within the browser display area, or it may invoke additional software such as a "plug-in" or external "helper application" to handle the file. Common file extension names include: ·

    - html, htm - Hypertext Mark-up Language (a web page) ·

    - gif, jpeg, tiff - A picture ·

    - wav, au, aif - A sound file ·

    - mov, avi, mpg, qt - a video clip ·

    - exe - An executable program for DOS/Windows (might be a self extracting archive) ·

- zip - A compressed archive based on PKzip (commonly used for DOS/Windows) ·

- hqx - A file that has been "binhexed" (commonly used for Macintosh archives)

Importance of Reading URL's. As you access online information resources from all over the Internet, it is important to read the URL of the displayed information (This information is displayed near the top of the web browser). The URL may be able to help you judge the value of the information. Consider the following examples:

1. Which web page would you use to influence your business unit's strategic plan?

   - http://www.sec.gov/rulings/exchanges/regulations.html

   - http://party.college.edu/freshman/joe/trading_regs.html

2. Which software program would you like to download?

   - ftp://ftp.microsoft.com/software/patches/fixit.exe

   - ftp://ftp.hackers.com/do/you/feel/lucky/trustme.exe

3. Which site is the US White House's official web site?

   - http://www.whitehouse.com

   - http://www.whitehouse.gov

The first two examples contain fictional URL's, which clearly illustrate the variety of resources you may encounter. The last example contains actual URL's, which show that URL's are not always a guarantee of authenticity. The first site is actually that of a commercial pornographic web site whereas the second is an official US government web site.
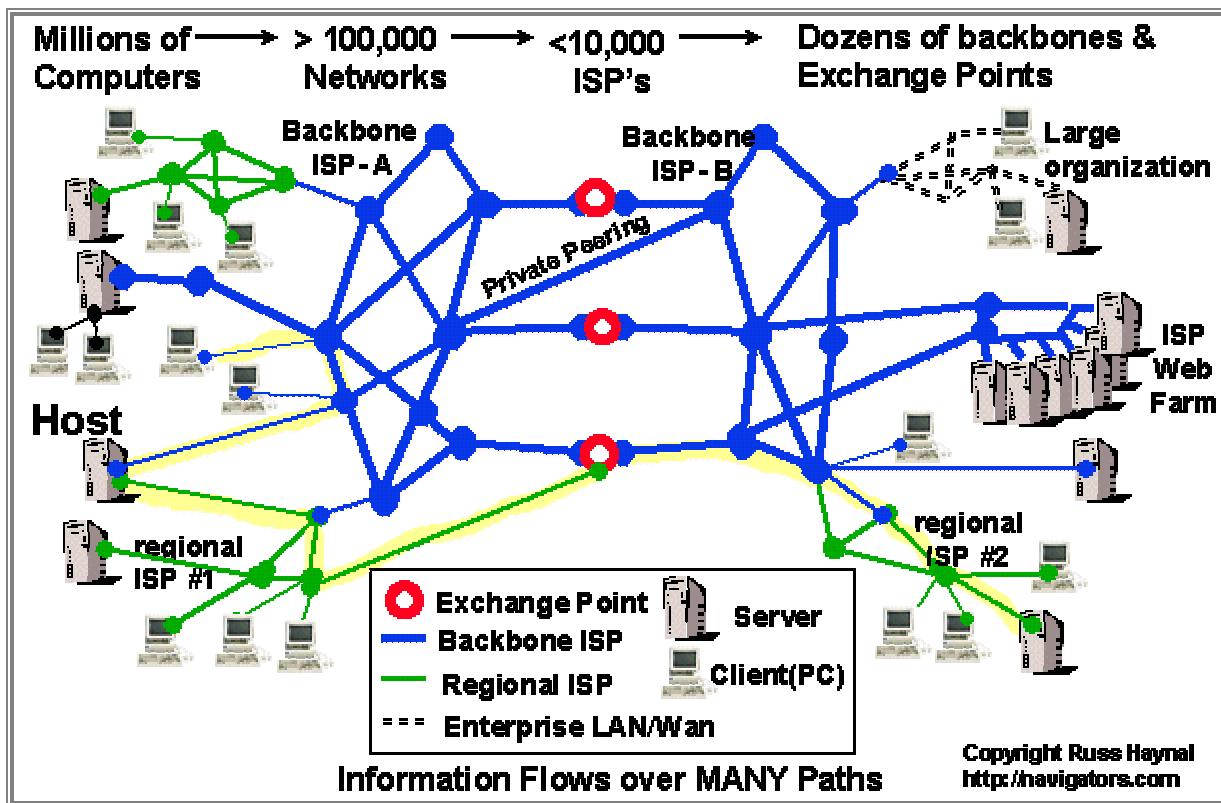
Tip: You should get into the habit of reading the URL of every web page before you even glance at the web page. URL's can also be used to decide which hyperlinks to select. While positioning your cursor over a hyperlink, the web browser will display (in the feedback area) the URL associated with the Hyperlink. This is a good way to "look before you leap"

Source: http://navigators.com/url.html

Traceroute

Using the *whois* function, you can determine the registered owner of a domain name. However, knowing who owns the domain may not satisfy your information requirement as ownership of a domain name is often less important than *where* the web server is physically located and *how* is it connected to the Internet. There is a network utility

called *traceroute*, which is often used to troubleshoot network connections. In a Unix or Windows environment, *traceroute* can be used to determine the specific network route taken from your workstation to reach a specific remote host. (Dos command: **tracertsitename.com**) Even more useful are the many Unix systems on the Internet that allow you to originate a traceroute from their location to any other location that you specify. Look at the following example of a *traceroute*.

Perform a traceroute to the host name

Look at this sample traceroute from http://www.boardwatch.com/to http://www.whitehouse.gov/:

1. t1 (204.144.169.2) 1.499 ms

2. t1.esoft.com (199.45.143.14) 9.549 ms

3. border-from-14-esoft.boulder.co.coop.net (199.45.130.13) 15.155 ms

4. core-gw-eth-0-2.boulder.co.coop.net (199.45.132.33) 33.277 ms

5. denver-cr2.bbnplanet.net (4.0.212.249) 19.309 ms

6. denver-br1.bbnplanet.net (4.0.52.1) 19.306 ms

7. oakland-br1.bbnplanet.net (4.0.1.133) 48.046 ms

8. oakland-br2.bbnplanet.net (4.0.1.78) 48.913 ms

9. sanjose1-br1.bbnplanet.net (4.0.1.74) 53.034 ms

10. mae-west2.us.psi.net (198.32.184.23) 47.213 ms

11. se.sc.psi.net (38.1.3.5) 148.414 ms

12. rc5.southeast.us.psi.net (38.1.25.5) 127.893 ms

13. ip2.ci3.herndon.va.us.psi.net (38.25.11.2) 139.433 ms

14. 198.137.240.33 (198.137.240.33) 134.218 ms

15. www2.whitehouse.gov (198.137.240.92) 141.93 ms

In the example, *boardwatch* gets its connectivity from "esoft.com" who get its connectivity from "coop.net" who is connected to BBNplanet (a backbone provider). BBNplanet Inter-connects with another backbone provider (PSI) through the MAE-West Connection point. It then appears that the whitehouse.gov web site is connected though PSI near Virginia. Recognize that an organization's website may not be located at the organization. The organization's website may be hosted someplace else. A more accurate approach to determine the location of the organization might be to do a traceroute to the organization's mailhost or proxy server.

Helpful information in reading *traceroute* results:

- Each Internet provider has their own naming convention. Most *traceroutes* travel across the largest Internet providers. A list of these as well as other helpful ISP-related information can be found on Russ Haynal's ISP page - http://navigators.com/isp.html.

- ISP node names may include an exchange point (MAE, NAP, PAIX). A list of the major exchange points can also be found on Haynal's ISP page - http://navigators.com/isp.html - naps

- ISP node names may describe the infrastructure topology (T1, T3, FDDI, ATM, HSSI, ETH)

- ISP node names may be the 3-letter code of the nearest airport in their node names. Here a list of all airport and city codes can be found at the following site - http://www.airportcitycodes.com/aaa/CCDBFrame.html.

- ISP node names may be just an IP Number. If you are trying to "read" the IP numbers, you should first understand how IP addressing works, how IP numbers are assigned, and then browse an IP network index. See the Internet Assigned Numbers Authority (http://www.iana.org) for information on IP addressing. There are several IP network indexes, the IP Network Index hosted by

dragonstar.net is one such IP network index -
http://www.crackinguniversity2000.it/Ipindex/
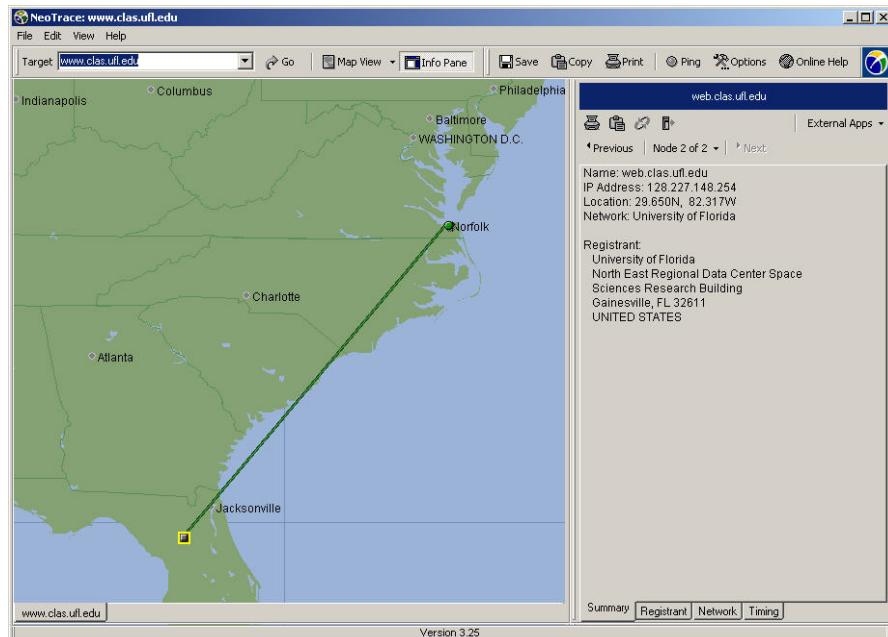
Traceroute Utilities:

You can perform *traceroutes* from various sites including: http://www.traceroute.org, http://www.geektools.com/traceroute.html, the Club Traceroute web page contains links to many *traceroute* resources and is found at http://www.amazing.com/Internet/club-traceroute.html.

Source: http://navigators.com/traceroute.html.

McAfee Visual Trace

A very convenient *traceroute* utility. McAfee acquired NeoTrace (shown below) and began to market an updated version under its new name.   Visual Trace displays the traceroute nodes as symbols with country flags. Visual Trace also associates the *Whois* for each network node in the trace. More information at http://www.mcafee.com/myapps/mvt.

You may also want to download *Visual Route*, which can overlay your traced route over a geographic map (http://www.visualroute.com/). A Java-based version of Visual Route is available online through this web page: http://visualroute.datametrics.com/

Another useful utility is the Hostname to Latitude/Longitude Converter which can be accessed through this web page: http://cello.cs.uiuc.edu/cgi-bin/slamm/ip2ll/. This utility not only provides the Latitude and Longitude for the server in question, but it also plots the location on a map.

Traceroute is only one, albeit important tool in analysing the source of information found on the Internet. One final point to remember is that it is important to realize that a website can be hosted anywhere despite where it might logically seem to be located. Organizations make the decision as to where they want to host a web site. Sometimes they have the resources and personnel to host the web site at their own facility, but it is also common for an organization to host their site at a commercial web hosting facility. For example, the Tico Times is a newspaper from Costa Rica (http://www.ticotimes.co.cr/), yet a traceroute reveals that the web site is physically hosted in Pittsburgh, Pennsylvania.

Section B: Evaluation Checklists

The following section contains a series of checklists, which can be used in the evaluation of various types of web pages. Virtually all web sources can be categorized under the following headings:

- Advocacy
- Business/Marketing
- News
- Information
- Personal

While not intended to be exhaustive, they provide a useful framework when evaluating web pages. Readers are encouraged to add their own steps based on their experience as they see fit. Note that in all of these cases the greater number of questions listed below answered *"yes"*, the more likely one is able to determine whether the source is of high information quality.

Evaluation Checklist for an Advocacy Web Page

An **Advocacy Web Page** is one sponsored by an organization attempting to influence public opinion (that is, one trying to sell ideas). The URL address of the page frequently ends in **.org** (organization).
Examples: Bellona Foundation (http://www.bellona.org/), Human Rights Watch (http://www.humanrightswatch.org/), and Amnesty International (http://www.amnesty.org/).

Criterion #1: AUTHORITY

1. **Is it clear what organization is responsible for the contents of the page?**

2. Is there a link to a page describing the goals of the organization?

3. **Is there a way of verifying the legitimacy of this organization? That is, is there a phone number or postal address to contact for more information? (Simply an email address is not enough.)**

4. Is there a statement that the content of the page has the official approval of the organization?

5. Is it clear whether this is a page from the national or local chapter of the organization?

6. Is there a statement giving the organization's name as copyright holder?

Criterion #2: ACCURACY

1. Are the sources for any factual information clearly listed so they can be verified in another source? (If not, the page may still be useful to you as an example of the ideas of the organization, but it is not useful as a source of factual information).

2. Is the information free of grammatical, spelling, and typographical errors? (These kinds of errors not only indicate a lack of quality control, but can actually produce inaccuracies in information.)

Criterion #3: OBJECTIVITY

1. Are the organization's biases clearly stated?

2. If there is any advertising on the page, is it clearly differentiated from the informational content?

Criterion #4: CURRENCY

1. Are there dates on the page to indicate:
   a. When the page was written?
   b. When the page was first placed on the web?
   c. When the page was last revised?

2. Are there any other indications that the material is kept current?

Criterion #5: COVERAGE

1. Is there an indication that the page has been completed, and is not still under construction?

2. Is it clear what topics the page intends to address?

3. Does the page succeed in addressing these topics, or has something significant been left out?

4. Is the point of view of the organization presented in a clear manner with its arguments well supported?

Source: http://www2.widener.edu/Wolfgram-Memorial-Library/webevaluation/advoc.htm

Evaluation Checklist for a Business/Marketing Web Page

A **Business/Marketing Web Page** is one sponsored by a commercial enterprise (usually it is a page trying to promote or sell products). The URL address of the page frequently ends in **.com** (commercial). **Examples:** Microsoft Inc. (http://www.microsoft.com/), DaimlerChrysler AG (http://www.daimlerchrysler.com/), as well as millions of other large and small companies using the web for business purposes.

Criterion #1: AUTHORITY

1. **Is it clear what company is responsible for the contents of the page?**

2. Is there a link to a page describing the nature of the company, who owns the company, and the types of products the company sells?

3. **Is there a way of verifying the legitimacy of this company? That is, is there a phone number or postal address to contact for more information? (Simply an email address is not enough.)**

4. Is there a way of determining the stability of this company?

5. Is there a statement that the content of the page has the official approval of the company?

6. Is there a statement giving the company's name as copyright holder?

Criterion #2: ACCURACY

1. Has the company provided a link to outside sources such as product reviews or reports filed with the SEC (the Securities and Exchange Commission) which can be used to verify company claims?

2. Are the sources for any factual information clearly listed so they can be verified in another source?

3. Is the information free of grammatical, spelling, and typographical errors? (These kinds of errors not only indicate a lack of quality control, but can actually produce inaccuracies in information.)

Criterion #3: OBJECTIVITY

1. For any given piece of information, is it clear what the company's motivation is for providing it?

2. If there is any advertising on the page, is it clearly differentiated from the informational content?

Criterion #4: CURRENCY

1. Are there dates on the page to indicate:
    a. When the page was written?
    b. When the page was first placed on the web?
    c. When the page was last revised?

2. Are there any other indications that the material is kept current?

3. For financial information, is there an indication it was filed with the SEC and is the filing date listed? For material from the company's annual report, is the date of the report listed?

Criterion #5: COVERAGE

1. Is there an indication that the page has been completed, and is not still under construction?

2. If describing a product, does the page include an adequately detailed description of the product?

3. Are all of the company's products described with an adequate level of detail?

4. Is the same level of information provided for all sections or divisions of the company?

Evaluation Checklist for a News Web Page

A **News Web Page** is one whose primary purpose is to provide extremely current information. Many international news outlets have web pages whose URL address of the page ends in **.com** (commercial), though many national news outlets web pages end in their respective nation's country code. One excellent example of this is the web page for BBC News – http://news.bbc.co.uk/. Other examples of news web pages are CNN (http://www.cnn.com/), Reuters (http://www.reuters.com/), and Agence France-Presse (http://www.afp.com/).

Criterion #1: AUTHORITY

1. **Is it clear what company or individual is responsible for the contents of the page?**
2. Is there a link to a page describing the goals of the company?
3. **Is there a way of verifying the legitimacy of the company? That is, is there a phone number or postal address to contact for more information? (Simply an email address is not enough.)**
4. Is there a non-web equivalent version of this material which would provide a way of verifying its legitimacy?
5. If the page contains an individual article, do you know who wrote the article and his or her qualifications for writing on this topic?
6. Is it clear who is ultimately responsible for the content of the material?
7. Is there a statement giving the company's name as copyright holder?

Criterion #2: ACCURACY

1. Are sources for factual information clearly listed so they can be verified in another source?
2. Are there editors monitoring the accuracy of the information being published?
3. Is the information free of grammatical, spelling, and typographical errors? (These kinds of errors not only indicate a lack of quality control, but can actually produce inaccuracies in information.)

Criterion #3: OBJECTIVITY

1. Is the informational content clearly separated from the advertising and opinion content?
2. Are the editorials and opinion pieces clearly labelled?

Criterion #4: CURRENCY

1. Is there a link to an informational page which describes how frequently the material is updated?
2. Is there an indication of when the page was last updated?
3. Is there a date on the page to indicate when the page was placed on the web?
   a. If a newspaper, does it indicate what edition of the paper the page belongs to?
   b. If a broadcast, does it indicate the date and time the information on the page was originally broadcast?

Criterion #5: COVERAGE

1. Is there a link to an informational page which describes the coverage of the source?
2. If you are evaluating a newspaper page and there is a print equivalent, is there an indication of whether the Web coverage is more or less extensive than the print version?

Evaluation Checklist for an Informational Web Page

An **Informational Web Page** is one whose purpose is to present factual information. The URL Address frequently ends in **.edu** or **.gov**, as many of these pages are sponsored by educational institutions or government agencies. **Examples:** dictionaries, thesauri, transportation schedules, calendars of events, and statistical data.

Criterion #1: AUTHORITY

1. **Is it clear who is responsible for the contents of the page?**
2. Is there a link to a page describing the purpose of the sponsoring organization?
3. **Is there a way of verifying the legitimacy of the page's sponsor? That is, is there a phone number or postal address to contact for more information?**
4. Is it clear who wrote the material and are the author's qualifications for writing on this topic clearly stated?
5. If the material is protected by copyright, is the name of the copyright holder given?

Criterion #2: ACCURACY

1. Are the sources for any factual information clearly listed so they can be verified in another source?
2. Is the information free of grammatical, spelling, and typographical errors? (These kinds of errors not only indicate a lack of quality control, but also can actually produce inaccuracies in information.)
3. Is it clear who has the ultimate responsibility for the accuracy of the content of the material?
4. If there are charts and/or graphs containing statistical data, are the charts and/or graphs clearly labelled and easy to read?

Criterion #3: OBJECTIVITY

1. Is the information provided as a public service?
2. Is the information free of advertising?
3. If there is any advertising on the page, is it clearly differentiated from the informational content?

Criterion #4: CURRENCY

1. Are there dates on the page to indicate:
    a. When the page was written?
    b. When the page was first placed on the web?
    c. When the page was last revised?
2. Are there any other indications that the material is kept current?
3. If material is presented in graphs and/or charts, is it clearly stated when the data was gathered?
4. If the information is published in different editions, is it clearly labelled what edition the page is from?

Criterion #5: COVERAGE

1. Is there an indication that the page has been completed, and is not still under construction?
2. Is there a clear indication of whether the entire work is available on the web or only parts of it?

Evaluation Checklist for a Personal Web Page

A **Personal Web Page** is one published by an individual who may or may not be affiliated with a larger institution. Although the URL address of the page may have a variety of endings (e.g. *.com, .edu*, etc.), a tilde (~) is frequently embedded somewhere in the URL. Some examples of personal web pages include Gary Capeci's web page – (http://www.gangland.com/), Julie's Astronomy Page (http://www.geocities.com/Astronomy30/), and The Missile Index (http://www.index.ne.jp/missile_e/index.html).

Criterion #1: AUTHORITY

1. **Is it clear what individual is responsible for the contents of the page?**

2. Does the individual responsible for the page indicate his or her qualifications for writing on this topic?

3. **Is there a way of verifying the legitimacy of this individual?** (Because it is difficult to verify the legitimacy of an individual, personal home pages may be a useful source for personal opinion but use extreme caution when using them as a source for factual information.)

Criterion #2: ACCURACY

1. Are the sources for any factual information clearly listed so they can be verified in another source? (If not, the page may still be useful to you as an example of the ideas of the individual, but it is not useful as a source of factual information.)

2. Is the information free of grammatical, spelling, and typographical errors? (These kinds of errors not only indicate a lack of quality control, but also can actually produce inaccuracies in information.)

Criterion #3: OBJECTIVITY

- Are the person's biases clearly stated?

Criterion #4: CURRENCY

1. Are there dates on the page to indicate:

    a. When the page was written?

    b. When the page was first placed on the web?

    c. When the page was last revised?

2. Are there any other indications that the material is kept current?

Criterion #5: COVERAGE

1. Is there an indication that the page has been completed, and is not still under construction?

2. If there is a print equivalent to the web page, is there a clear indication of whether the entire work is available on the web or only parts of it?

Source: http://www2.widener.edu/Wolfgram-Memorial-Library/webevaluation/perspg.htm

Section C: Validated Source Lists

Considerable effort is required to validate an Internet source. Industry leaders, once discovered and validated, may be realistically relied upon to maintain the same degree of accuracy and consistency in reporting. Other collateral sources must be continually validated to ensure their accuracy.

Many Internet users either fail to make the effort to validate information sources prior to using the information that they provide, or do not take steps to capture the source validation data that the evaluation process produces. Capturing this information allows the analyst to refer back to the source evaluation prior to returning to that source. Perhaps more importantly, it allows the analyst to share his sources with other collaborating analysts. Rather than simply providing a list of Internet hyperlinks, the inclusion of source evaluation data provides a broader understanding of the analytical judgements made in including a particular source while excluding another.

A number of options for maintaining and exchanging sources and source validation exist. A live web-page with hyperlinks directly to the source is an effective means with which to disseminate this data. Security can be achieved if this list is maintained behind a firewall or other protection protocol. Some organizations have used lists of sources arranged by subject and managed either in a spreadsheet or a text document.

Powermarks 3.5

The bookmark management program *Powermarks 3.5*[6] is a highly effective means with which to organize large amounts of Internet bookmarks as well as capture and maintain source evaluation data.
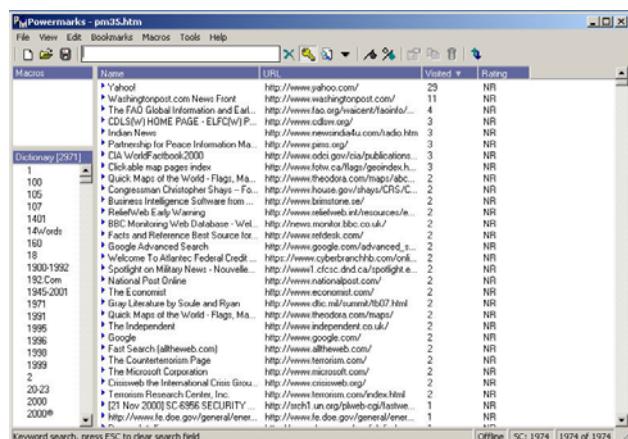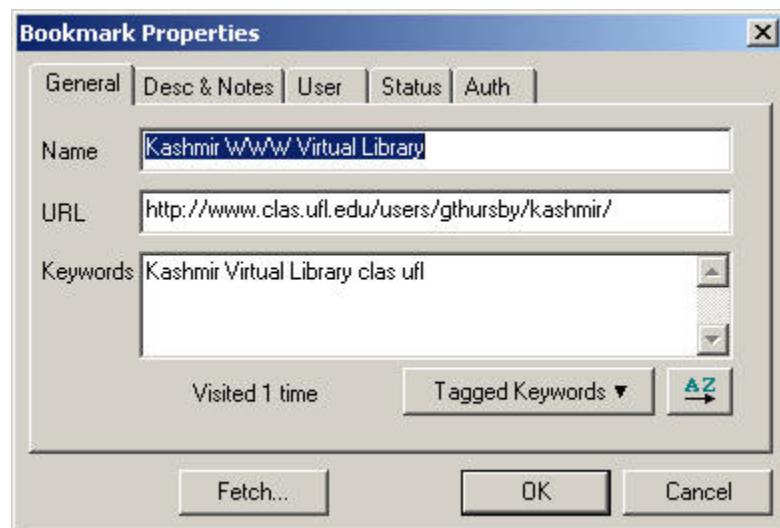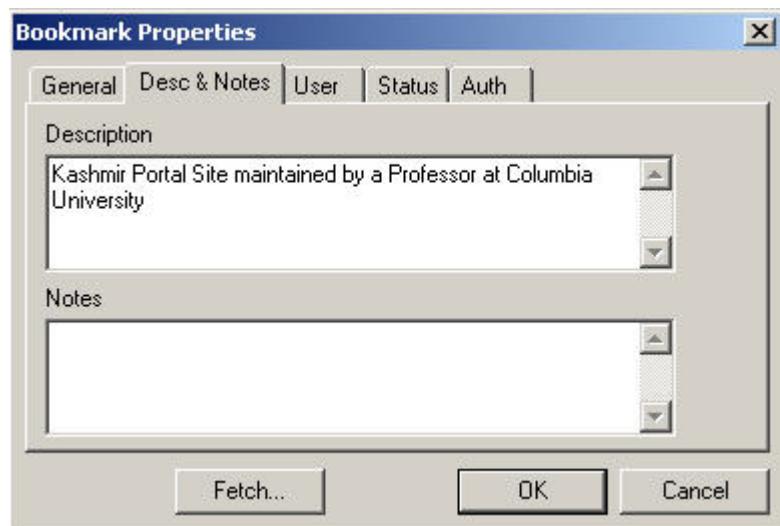

Figure 1 – Powermarks 3.5 Main Screen

---

[6] http://www.kaylon.com/power.html
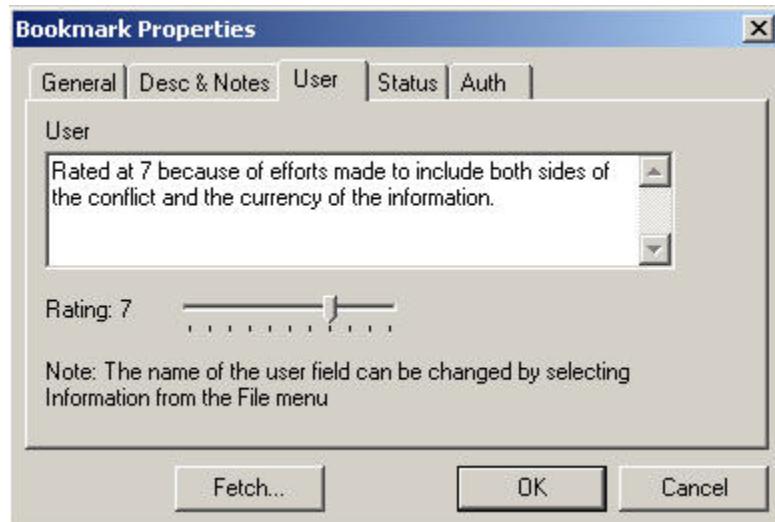
The traditional method of storing bookmarks has been in a hierarchical file structure by subject type. The main limitation of this method is that a particular link may have applicability across a number of subjects. The only means with which to provide that link with visibility across these topics was to store it in several places. Contrasting with that approach, Powermarks stores bookmarks in a flat file database that is searched by keyword.

When a link is saved, the properties screen opens and the analysts are able to list in the keywords field, all the relevant subjects related to that particular page. When retrieving a link, the Powermarks main screen opens with a search box that asks for keywords to be entered to retrieve links saved in the local database. This is a highly effective means with which to manage links that cut across traditional subject lines.

The description and notes field allow the analyst to identify the source in plain text with amplifying details. Rather than a link to the Internet, the source description can provide a sense of why it is important to an analyst.

The user-rating field provides the analyst with the means with which to attach a subjective rating of 1-10 to the site and the information that it provides. It also provides a text box to explain why the rating has been attached to the source.

Sources of known reliability can be used to corroborate new information sources. Source reliability is not static. It is affected by new biases, changes in funding sources, and personnel changes, among other factors. Sources should always be monitored closely for shifts in their reporting. The maintenance of accurate records is essential for tracking source consistency and identifying and monitoring new sources.

By maintaining an accurate source list, time can be saved in the collection of information from the Internet. While a number of methods for the maintenance of these records are available, commercially available solutions like Powermarks provide an effective, low-cost solution to the record maintenance problem.

While a server-based solution is not available for this particular product, it does have the means with which to export and import Powermarks files and to maintain its database on its corporate server for collaboration purposes.

Section D: Effective Summary

The combination of information browsers, word processors and the myriad information sources available via the Internet produces a highly tempting option for analysts to gather vast quantities of information on topics of interest. Even with disciplined collection strategies, information overload is an ever-present problem. While open sources can contribute to satisfying an information requirement in a timely fashion, the resulting information must be structured so as to be relevant to the analyst and easy to use and understand for others who receive it.

Effective summary is an important yet elementary skill necessary to shape large quantities of information into a manageable form. When gathering data for further dissemination, the inclusion of a summary paragraph that captures the essence of the information can be the difference between whether the information is reviewed or rejected.

Copernic Summarizer

The most effective means of summary is the application of analytical skill to information in order to capture the salient points relating to the stated information requirement. A number of commercial products are now being marketed to assist in the rapid summary of information sources. One is *Copernic Summarizer*[7].
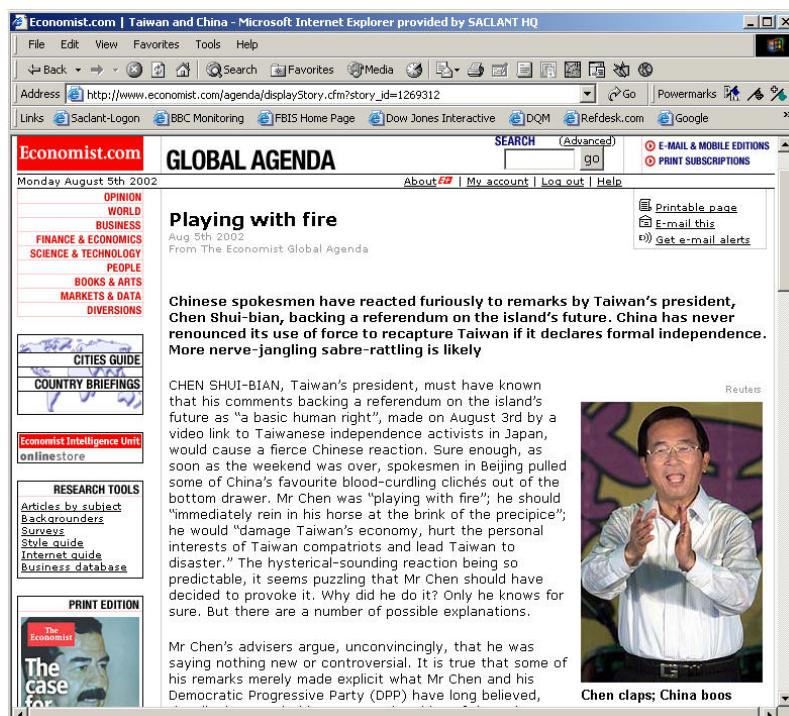

Figure 2 – Complete Web-page 1200 words in length

---

[7] http://www.copernic.com/products/summarizer/

Copernic Summarizer can analyse text of any length, on any subject, in any one of four languages (English, French, German and Spanish), and create a summary as short or as long as you want it to be. It can summarize Word documents, Web pages, PDF files, e-mail messages and even text from the Clipboard.
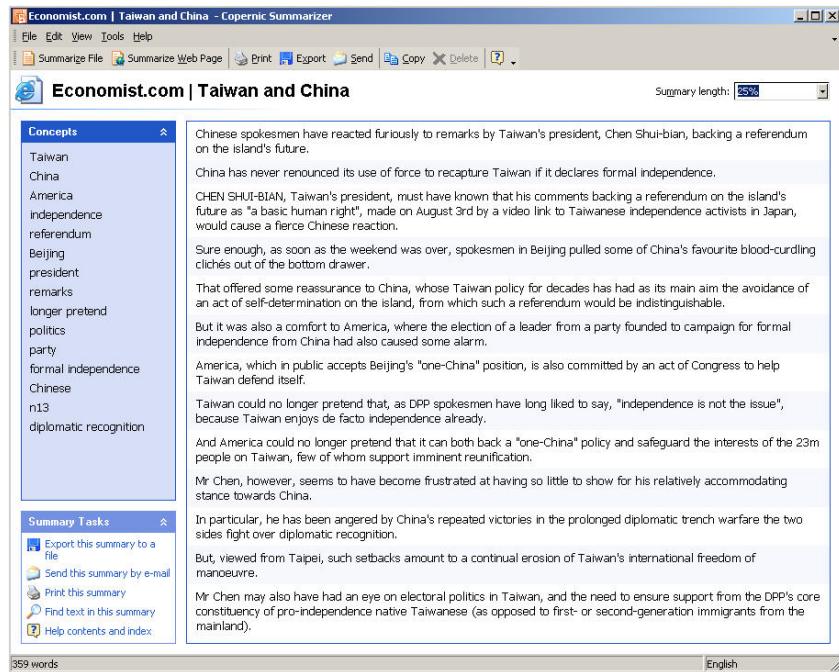

Figure 3 – Copernic Summarizer produced summary of previous article (25% of original length).

Summary reports include key concepts and key sentences according to the configuration of the software. It is possible to summarize a whole document or a selected part of the document. If the summary does not capture the essence of the topic, it is possible to refine the summary report by deleting selected concepts and sentences, with an automatic summary updating when a concept is deleted.

The resulting summary report may be exported in various file formats (HTML, XML, Rich Text Format and Text file) and appended to files. These files can be distributed by the integrated email function, printed, or copied into other documents.

While not a solution to the problem of information overload, the judicious use of these technologies can assist in the reduction of large documents to more manageable lengths.

CHAPTER V: DISSEMINATION

The essential fundamentals of effective dissemination are four-fold: the right person must receive the right product at the right time and in the right format.   While this principle is well accepted in classified intelligence disciplines, it is no less relevant for Internet-derived material. Given the ease with which material can be collected and collated from the Internet, its dissemination requires probably even greater care if it is not to further contribute to information overload.

Information provided must not only be accurate and relevant, but it must be structured and presented in such a way that the recipient can use it easily, assimilate it quickly, and understand its significance in time to act on it effectively. Since each level of command both within and outside an organization have differing information-detail and time-management requirements, intelligence and information support products must be structured, presented, and disseminated in ways that fit these varying requirements.

Section A: Report Layering

An effective means with which to present collated information is with a "layered" product. Through the use of an Executive Summary and a Table Of Contents, followed by only the relevant material, a product collected from Internet sources can simultaneously satisfy the information needs of a variety of users.

Special Report: J&K Elections And Beyond

Executive Summary

1. **Assessment**: The 16 Sep – 8 Oct Jammu & Kashmir (J&K) elections have two major implications: one, they will not address Kashmiri aspirations for self-determination, and, two, militant violence and low voter turnout will probably return India-Pakistan tensions to pre-June levels, leading to an increased risk of war from mid-October to December [before winter takes hold in Kashmir]. India's goal is to make this round of elections at least appear more representative of Kashmiri interests, in order to legitimize New Delhi's rule in J&K. Pakistan wants to avoid successful J&K elections, since it would weaken Islamabad's position on Kashmir, and may encourage militant disruption of the polls. Islamabad may be over-confident in assuming that nuclear deterrence would discourage Indian military action. Washington's challenge will be to respond to the elections in a way that is perceived as impartial, and to create incentives for both sides to examine other non-military approaches to Kashmir. [For more, see In-Depth Analysis].

2. **Background**: India has held elections for the J&K Legislative Assembly since 1950. The current Assembly's 6-year term expired in 2002. Kashmiris have traditionally complained of widespread fraud, abuse by Indian security forces, and ballot-box stuffing during Assembly polls. There is independent evidence from Indian human rights groups, journalists, and international academics to support these claims. India cites Pakistani-supported 'subversion' as a major reason for past election failures. The prospect of 'subversion' vastly increased since the onset of the [initially indigenous] Kashmiri insurgency in 1989. Pakistan's interest in weakening Indian control in J&K led it to provide material support and training to the Kashmiri insurgents. In the early 1990s, several new Pakistan-based Kashmiri militant groups were formed by Afghan war veterans, and this helped transform [although it was not the only factor] the insurgency's identity from secular nationalist to a transnational religious movement. During this time, the Kashmiri quest for self-determination lost its distinction from Islamic extremism. This laid the basis for New Delhi's current characterization of Kashmiri militants as foreign terrorists in a 'proxy war' by Pakistan.   [For more, see Background].

3. **Discussion**: India will conduct the September/October elections under foreign media and diplomatic scrutiny and with significant international expectations. The polls will be held in four phases on September 16, September 24, October 1, and October 8, in various districts of J&K. India has increased security in J&K for the upcoming elections, in hopes of preventing large-scale militant attacks. Police, para-military, and Army units will be stationed around polling booths. Despite heightened security measures, there are daily incidents of violence against election candidates and a general sense of fear amongst election workers, politicians, and the Kashmiri population. Diplomats from the US, UK, Australia, Canada, Denmark, France, Germany, Greece, Italy, Japan, Luxemburg, Netherlands, Spain, Sweden, Switzerland will observe the elections, although there will be no formal international monitoring. Results are expected to be announced on 12 October. [For more, see Discussion].

Special Report:
J&K Elections And Beyond

Table of Contents

An approach in structuring information is to consider how those who receive it use information. A safe rule-of-thumb is the following:

- Executives (decision-makers) require executive summaries
- Staff officers require relatively complete summaries focused on their assigned areas with enough background for them to understand the "big picture"
- Analysts often need a detailed treatment of the subject within their assigned area of interest in order to make their own assessment of information reliability
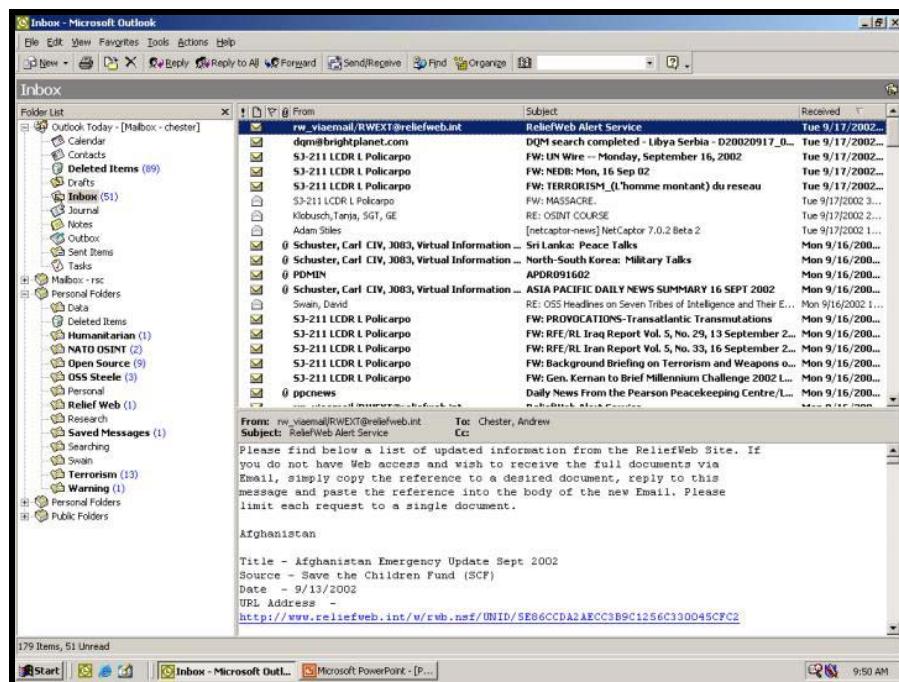
Section B: Dissemination with *Microsoft Outlook*

An alternate means of dissemination available to NATO Intelligence staffs is direct email of relevant collected material at the time of its collection. For particularly time-sensitive material, this may be appropriate. *Microsoft Outlook* is the email standard within NATO commands. This software supports distribution lists so that one researcher can collect material and disseminate it quickly to all staff members who are working on a subject. This is highly effective in an intelligence watch structure. A watch officer can be supported by dedicated open source analysts who are able to manage a mission-focused collection effort on their behalf.
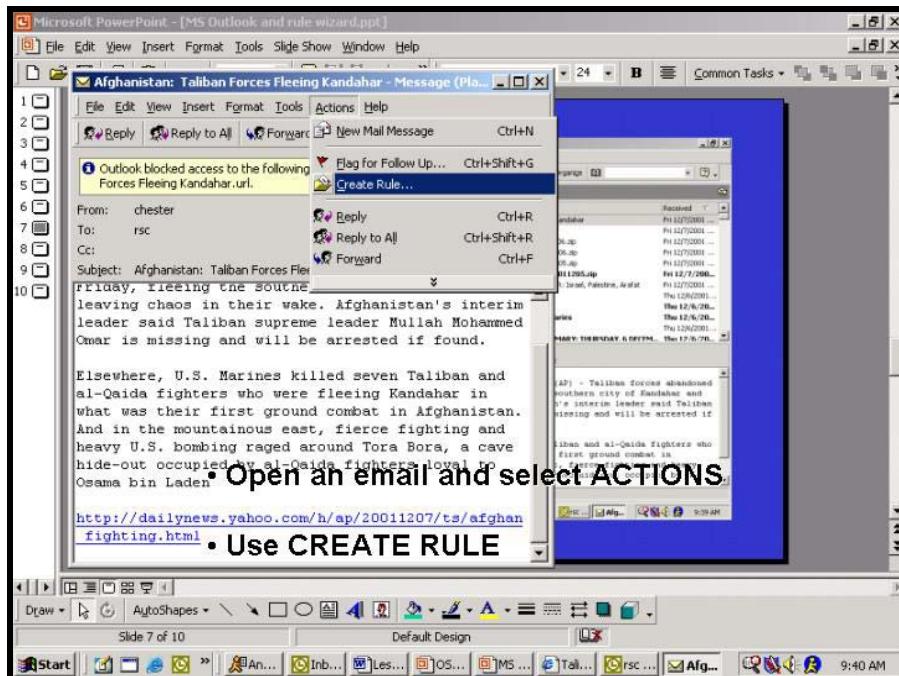
Some OSINT organizations disseminate vast amounts of material (several hundred messages daily) directly to a specific set of users through distribution lists. Unless the user is prepared to deal with a large flow of messages, the collected material quickly is treated as *SPAM* by its recipient. *Outlook* can help with this.

It is possible to use Outlook as a database for collected material placed either manually or automatically within separate folders. These folders can then be searched locally for collected material. Each user is able to establish a series of folders, which can be populated with relevant material collected from the Internet or sent to an email account by others.

Most *Outlook* users are familiar with folders; fewer understand how they can be automatically managed. The **rules** function in *Outlook* can be optimized to help mitigate the impact of a large volume of email.

The Rules Wizard can be used to automatically store files received within an email inbox based on criteria set by the user. Emails or copies of emails received from a particular email address or with a specific keyword in its title can automatically be placed into a file for later reference. This provides the recipient with the flexibility to either deal directly with the information as it is received or later when they are ready to conduct analysis on a specific topic.

- Open an email and select ACTIONS
- Use CREATE RULE

Section C: Dissemination and Classification

Intelligence is classified to protect sources, methods and intentions. While information hosted by Internet sources is inherently unclassified, its collection in support of some mission objectives can make its retention or dissemination by an intelligence staff a sensitive issue. It is established intelligence doctrine that the originator of the information is responsible for its classification. This goes for material collected from the Internet as well.

Dangers arise when information gathered from the Internet is classified without good reason. Over-classification of any intelligence inhibits its effective dissemination. Once material collected from Internet sources is given any classification, its dissemination can only be achieved via classified systems. In some cases, this may limit the utility of the information particularly if its collection was intended to support a broader international effort outside the traditional military structure within NATO.

When information gathered from the Internet is merely reproduced within a classified intelligence product, confidence in the classification system is compromised. If an intelligence analyst can gather a particular item from an Internet source, it is likely that other members of the staff, equally interested in the subject, will be familiar with the source. Seeing Internet material merely repeated back under a classified banner undermines the intelligence process.

While in some cases it makes sense to protect the source, in most cases it does not. Information gathered from the Internet should be treated in a manner appropriate to the mission. For example, general information collected on the subject of terrorism in southern Europe would not expose any NATO plans or operations. Public statements by both military and political leaders within the Alliance have made clear NATO's interest in the field of terrorism. Even a compilation of relevant documents is unlikely to require protection with classified markings. However, specific collection of information from the Internet on the presence of terrorist cells or particular leaders in the Naples, Italy area would require greater protection both in the manner in which the information was gathered and in the steps taken to protect it.

One of the principal advantages that OSINT brings to the intelligence process is the ability to rapidly prepare intelligence products that can be easily shared with coalition partners. The use of open sources to construct an intelligence picture that closely mirrors the classified intelligence picture but that does not compromise sources and methods provides a commander and his intelligence staff with an important tool to support coalition warfare. NATO operations have historically been conducted in a multinational environment with NATO forces alongside those of non-NATO troop contributing nations providing the military component within a larger international effort. Reliable and sharable unclassified information enables the commander to benefit from information exchanges outside of his command structure. These information exchanges foster a common view of the operating area among all mission elements and nurtures an environment of information sharing within a coalition.
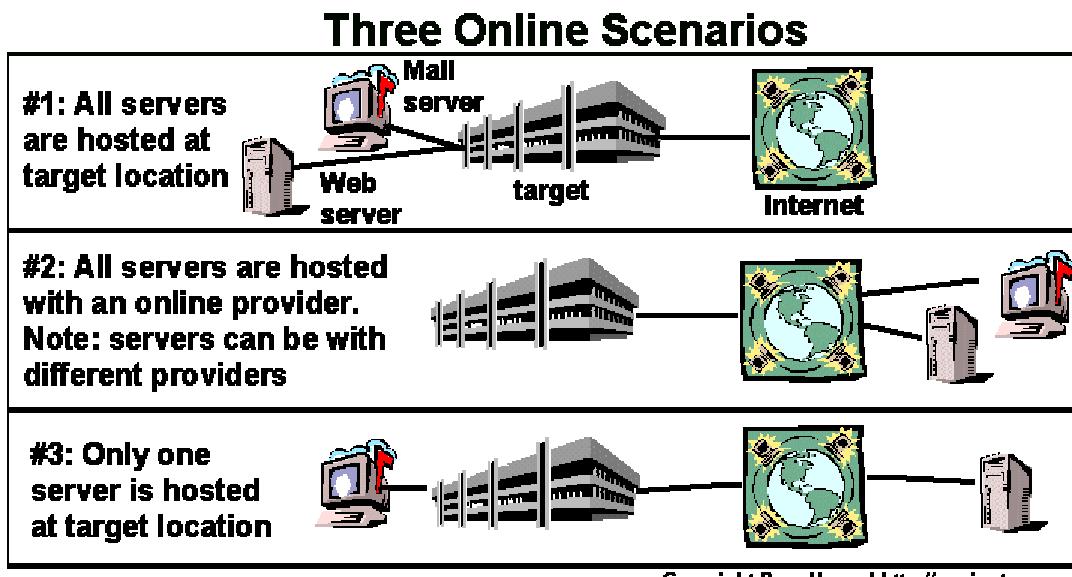
While there are always limitations to information sharing, at least a robust open source exploitation capability assures a minimal common appreciation of the situation including terrain and civil factors. This benefit crosses mission essential civil-military lines. However, a strict adherence to classifying all products that include intelligence staff in their preparation eliminates these potential advantages.

A final point on dissemination in a coalition environment is the concept of "bottom-up" report compilation. By beginning with unclassified sources, a report can be created on a subject that is instantly shareable with whomever the commander chooses. Given the broad range of information sources available today virtually any subject can initially be evaluated with open sources. After this initial process, relevant classified collateral reporting can be added to increase the intelligence value of the product. Ultimately, a variety of products can be created at various levels of classification to support mission requirements. This process contrasts sharply with the traditional approach of sanitizing/declassifying intelligence to support coalition partners. Yet, it is a faster process that both makes use of open sources to produce products for wide distribution and serves to integrate open source derived material directly into the classified intelligence production process.

Building intelligence products by the "bottom up" method provides a quick and efficient means of building products for multiple security domains and audiences without having to resort to a disclosure process.

ANNEXES:

Annex A: More information on Source Evaluation

Three Online Scenarios

#1: All servers are hosted at target location — Mail server, Web server, target, Internet

#2: All servers are hosted with an online provider. Note: servers can be with different providers

#3: Only one server is hosted at target location

Copyright Russ Haynal http://navigators.com

Consider these scenarios:

1. The organization hosts all their servers at their own facility. A traceroute to the web server or email server will also indicate how the entire organization is connected

2. The organization chooses to host all of their servers with various out-sourced solutions. Recognize that the web server, email server, and Access provider to the organization's building can all be with different vendors. The most assured way of determining how this organization is connected would be to do a traceroute to a connection persona of one of the employees who has visited a web site (This method works for all scenarios).

3. Some servers may be out-sourced, some may be located locally. A traceroute to a locally hosted server will show how the organization is connected

Problem - How do you discover the server locations associated with a domain name? The Information is located online in the DNS Records for the domain. Just "ask" any domain name server for the DNS record of the target site. This information is widely used to properly route web requests and email messages to the domains servers wherever they may be located.
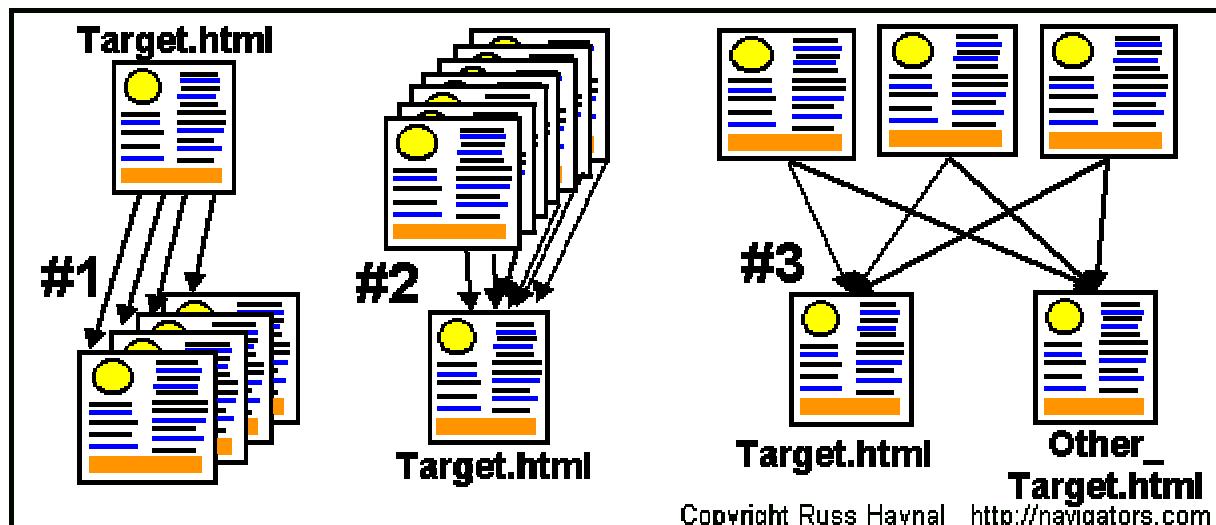
Here are several integrated multi-use tools:

- Network-tools.com (http://www.network-tools.com/) is another great resource (click on one of the numbers to go to a mirrored page) - Be sure to try **DNS Records**.

- Sam Spade (http://www.samspade.org/) tries the **scan rDNS** to scan a network block for named hosts.

- Codeflux (http://www.codeflux.com/tools) has easy interface to whois, dig, traceroute, etc.

- The multiple DNS look-up engine found here - http://www.bankes.com/nslookup.htm- allows you to enter a web domain, or IP number, and this site will allow you to discover who are the target's neighbours (from an IP number point of view)

- Zone Edit has a web-based DNS look-up that can be found here http://www.zoneedit.com/lookup.html

Searching Upstream

You've just found a great target web page. You've even followed the links "downstream" from that web page to other web pages.

Now do the most important search. Find the web pages that point **towards** the target web page (searching "upstream")
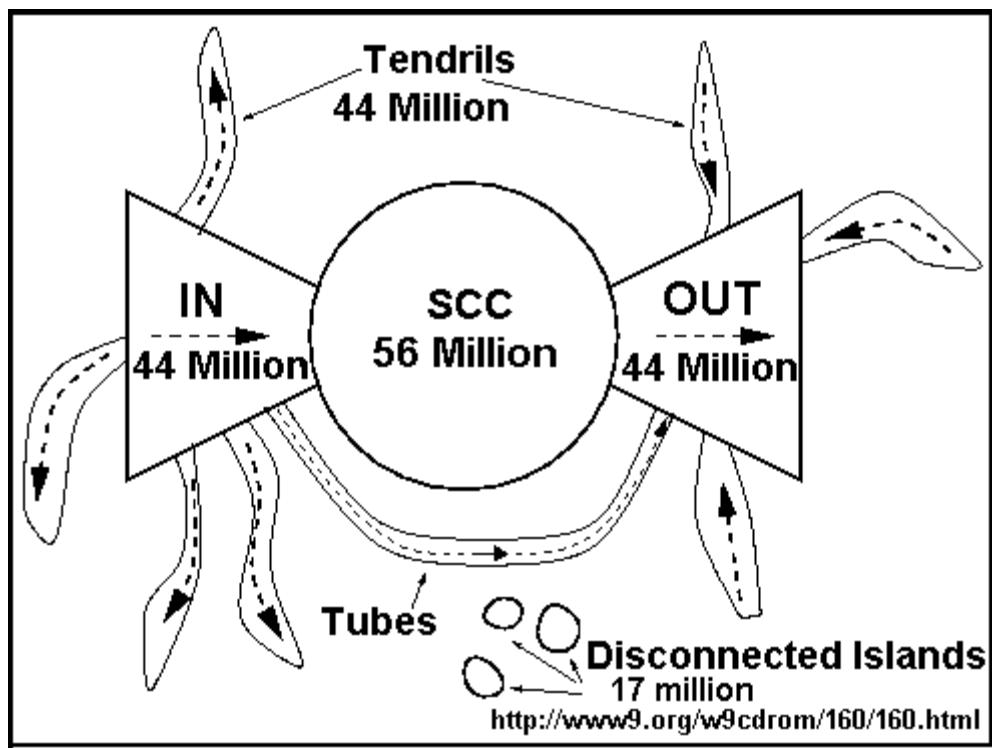

Look at these three scenarios and see the value in these different approaches to searching.

1. You discover a valuable web page called target.html. Most people simply explore the hyperlinks contained within target.html. Those links take you to places suggested by the author of target.html. In short, all you are able to discover are the pages which are located "downstream" of target.html

2. You discover a valuable web page called target.html. In order to help judge the importance, value, popularity of the page, it would be nice to know "How many other web pages contain hyperlinks pointing towards target.html. This shows how many web authors know about the target site, and felt that the site was "good enough" to deserve a hyperlink from their own web site." If a web page looks fairly anonymous, you might be able to infer something about the source of the web page based on who else links toward the web page. Fortunately some search engines (like AltaVista) will allow you to search for all web pages, which contain a hyperlink to a specific URL. This is what is called "searching upstream" of a web page. Here is an example search: the following search would show you what pages link to the White House web site which also have a domain name ending in .jp +link:whitehouse.gov +domain:.jp.   Note that some search engines can rank their search results based on the "popularity" of a web paged derived from how many links point toward that web page.

3. You discover two valuable web pages called target.html and other_target.html. If you can find any web pages that link to both of these target pages, then you may discover a great "directory" such as "Joe's list of targets on the Internet" for the subject covered by these target pages.

Here are additional search examples based on Altavista's format:

+link:resource.com	Web pages containing links toward resource.com
+link:resource.com/pageA.html	web pages containing links toward the specific web page
+link:resource1.com +link:resource2.com	Web pages which link towards both resources. This is a good way to discover virtual libraries
+link:newpaper_1.com +link:newspaper_2.com	Web pages which link to these two newspaper (might also link to many other newspapers)
+link:resource.com +domain:.ru	Web pages hosted on .ru machines which contain a hyperlink to resource.com
+link:resource.com +host:gov.ru	Web pages hosted on gov.ru machines which contain a hyperlink to resource.com
+search_term +host:gov.ru	Web pages hosted on gov.ru machines which mention search_term
+host:resource.com	Web pages from Resource.com - can be used to size a web site

Searching upstream is not only a good idea, it is actually **necessary** in order to even reach almost half of all web pages known to Alta Vista

Detailed research paper on this: http://www9.org/w9cdrom/160/160.html

Source: http://navigators.com/search_upstream.html

Annex B: Selected Information Sources

The following Internet resource pages were compiled by the Intelligence Staff of CINCEASTLANT. They provide both a resource of immediate value to analysts as well as some practical examples of how to employ successful Internet-based search strategies in support in a variety of intelligence problems. The date following the titles below represents the last time that the collection was updated.

a. Terrorist Threat Research (Rev 04 Mar 02)

b. Hostile Intelligence Threat Research (15 Feb 02)

c. Criminal Threat Research (15 Feb 02)

d. Medical Threat (Rev 08 May 02)

e. Geo-Political, Military & Country Information Research (Rev 04 Mar 02)

f. Geo-Political & Military Information – Russian Annex (Rev 04 Mar 02)

g. Gazetter, Port, Geodessy and Map Research (Rev 02 May 02)

a. Terrorist Threat Research (Rev 04 Mar 02)

BACKGROUND

Most often, research on terrorist threats will be incorporated into a Force Protection (FP) Threat Summary (THREATSUM) or Briefing. Both of these products should clearly and concisely define the actual and/or potential dangers to a specific allied/coalition force, in a specific geographic area of operations during a specific period of time. ***In all cases, the relevant "raw" research documents should be saved to an appropriate file folder under the country or area of concern.***

GUIDELINES

Confirm or refute the existence of a threat. Traveler warnings and terrorism reports issued by the State Departments or Foreign Ministries of the western governments are probably the best source for general force protection information. These documents provide credible general and specific information about terrorist threats as well as short-term and/or trans-national conditions posing a significant risk to the physical security of Allied/Coalition Force.

- U.S. Department of State Travel Warnings
  (http://travel.state.gov/travel_warnings. html).

- U.S. Department of State, Crime Control Fact Sheets
  (http://www.state.gov/g/inl/crm/fs)

- U.S. National Security Council, International Crime Threat Assessment
  (http://clinton4.nara.gov/WH/EOP/NSC/html/documents/pub45270/pub45270 chap3.html

- Canadian Department of Foreign Affairs, Travel Advisories
  (http://voyage.dfait-maeci.gc.ca/destinations/menue.htm).

- British Foreign and Commonwealth Office travel advice notices
  (http://193.114. 50.10 /travel/countryadvice.asp).

- U.S. Navy, Office of Naval Intelligence (ONI), Worldwide Threat to Shipping
  (http://pollux.nss.nima.mil/onit/onit_j_main.html)

- Center for Defense Information (http://www.cdi.org) - This site has weekly articles compiled or produced by academics which include information on world terrorist, weapons of mass destruction, military defense, etc.

Define the historical activities of the specific terrorist and subversive groups/ organizations residing and/or operating in, or transiting through the country or area of interest. The below listed databases provide historical data on a wide variety of well-known groups.

- U.S. Department of State Patterns of Global Terrorism Annual Reports (http:// www.state.gov/s/ct/rls/pgtrpt/) provide detailed assessments of foreign countries where significant terrorist acts occurred during the reporting period.

- American Federation of Scientists (FAS) Liberation Movements, Terrorist Organizations, Substance Cartels, and Other Para-State Entities (http://www.fas.org/ irp/ world/para/index.html) is a listing of 370 terrorist/subversive groups (many with hyperlinks to background information). Additionally, Search FAS (http://www.fas. org/) can be used to research the organization's databases.

- U.S. Navy Postgraduate School Terrorist Group Profiles (http://web.nps.navy.mil/ ~library/tgp/tgpndx.htm) is a listing of 370 terrorist/subversive groups (many with hyperlinks to background information).

- International Policy Institute for Counter-Terrorism (http://www.ict.org.il/) is an Israeli-based research database on terrorist organizations and events. Select the "Search Text" icon and enter the terrorist group name to access data.

- Emergency Response & Research Institute, Inc. (ERRI) Counter-Terrorism Archive (http://www.emergency.com/cntrterr.htm) provides information on World-Wide Terrorism Events, Groups, and Terrorist Strategies and Tactics.

- Jane's Information Group (http://www.janes.com/) is a **subscription service** providing access to highly credible data on geo-political, military, intelligence and terrorism issues. However, **free access is authorized to selected items** via Jane's Search (access through the homepage), the Sentinel Risk Pointers (http://www.janes.com/ regional_news/sentinel/risk_pointers.shtml) and International News Briefs (http://www.janes.com/regional_news/international/news_briefs/foreign_repor t_toc_2001-1.shtml).

- Economist Intelligence Unit (http://www.eiu.com) is a **subscription service** providing access to highly credible data on regional and national geo-political, economic and security issues. However, **free access is authorized to selected items** via the Data Services and ViewsWire web sites contain country specific historical and current political, economic and security information.

- EIU ViewsWire (http://www.viewswire.com/) is a daily country analysis service covering 195 countries and providing an unrivalled and timely database of background information, analysis and forecasts.

- Separatist, Para-military, Military, Intelligence, and Aid Organizations (http://members.home.net/bob.cromwell/security/Index.html#L) is a privately maintained home-page listing a large number of terrorist, subversive and/or "freedom fighter" organizations. Hyperlinks provide access to organization backgrounds.

- Political Terrorism Database (http://polisci.home.mindspring.com/ptd/) is a privately maintained database divided up into geographic areas containing an index to each region's terrorist groups as well as an international terrorism incident database.

- Terrorism Research Center, Calendar of Significant Events (http://www.terrorism. com/calendar/Calendar.html) is a compilation of dates that have importance for terrorism research and analysis.

- National Center for Policy Analysis (NCPA), Terrorism and Nuclear Proliferation archive of articles and press releases concerning the risk of nuclear terrorism (http:// www.ncpa.org/ pi/internat/intdex10.html).

Define current terrorist and subversive activity in the country or area of interest. The below listed databases, metasearch engines and news search engines/sites should be searched employing the search strings indicated below.

- U.S. Department of State, Overseas Security Advisory Council (OSAC) Daily Global News (http://www.ds-osac.org/globalnews/default.cfm) keyword search provides timely information on foreign events related to security.

- Vivisimo (http://www.vivisimo.com/) a high quality metasearch engine.

- Ixquick (http://www.ixquick.com/) a high quality metasearch engine supplementing Vivisimo.

- NewsTrawler (http://www.newstrawler.com/nt/nt_home.html) is a metasearch engine for news archives and articles from worldwide sources. Search categories are By Country, Across Countries, Category and Across Categories.

- The Ultimate Collection of News Links (http://pppp.net/links/news/) is a reference source of media hyperlinks by region and country. Conduct searches of the appropriate news media source(s) in the country or area of interest.

- News Index (http://www.newsindex.com) is a search engine for current news stories from worldwide sources.

- Pandia Newsfinder (http://www.pandia.com/news/index.html) is a search engine for current news and links to news sources.

- AlertNet (www.alertnet.org) is a Reuters-sponsored service providing global news, communications and logistics services to the international disaster relief community. The public pages, accessible to anyone, carry a live news feed and articles describing how relief agencies are responding to the latest humanitarian crises.

- BBC World Service (http://www.bbc.co.uk/worldservice/).

Use standard "search strings" employing simple Search Engine Math or Boolean Logic commands such as those listed below.

- *terrorist/subversive group name*
- *country name* **+terrorism**
- *country name* **+extremist***
- *country name* **+subversive**
- *country name* **+underground**
- *country name* **+bomb***
- *country name* **+explosion**
- *country name* **+assassin***
- *country name* **+arson**
- *country name* **+kidnap***
- *country name* **+demonstration**

b. *Hostile Intelligence Threat Research* *(15 Feb 02)*

BACKGROUND

Counterintelligence is a critical element of the Force Protection Threat Summary; however, the availability of credible open source databases and profiles on civilian and military intelligence services is extremely limited. To clearly define the actual and/or potential threat to allied/coalition forces, the analyst probably will have to rely heavily on news articles obtained from multi-national, regional and/or local media sources. ***In all cases, the relevant "raw" research documents should be saved to an appropriate file folder under the country or area of concern.***

GUIDELINES.

Identify the hostile intelligence service(s) residing and/or operating in the country or area of interest.

- American Federation of Scientists (FAS), World Intelligence and Security Agencies page (http://www.fas.org/irp/world/index.html) provides hyperlinks to profiles on intelligence agencies of selected countries.

- Jane's Information Group (http://www.janes.com/) is a **subscription service** providing access to highly credible data on geo-political, military, intelligence and terrorism issues. However, **free access is authorized to selected items** via Jane's Search (access through the homepage), the Sentinel Risk Pointers (http://www.janes.com/ regional_news/sentinel/risk_pointers.shtml) and International News Briefs (http://www.janes.com/regional_news/international/news_briefs/foreign_repor t_toc_2001-1.shtml).

- Economist Intelligence Unit (http://www.eiu.com) is a **subscription service** providing access to highly credible data on regional and national geo-political, economic and security issues. However, **free access is authorized to selected items** via the Data Services and ViewsWire icons.

Define historical and current hostile intelligence activities that may be directed against allied/coalition forces. The below listed databases, metasearch engines and news search engines/sites should be searched employing the search strings indicated in paragraph B.2.3.

- U.S. Department of State, Overseas Security Advisory Council (OSAC) Daily Global News (http://www.ds-osac.org/globalnews/default.cfm) keyword search provides timely information on foreign events related to security.

- Vivisimo (http://www.vivisimo.com/) a high quality metasearch engine.

- Ixquick (http://www.ixquick.com/) a high quality metasearch engine supplementing Vivisimo.

- NewsTrawler (http://www.newstrawler.com/nt/nt_home.html) is a metasearch engine for news archives and articles from worldwide sources. Search categories are By Country, Across Countries, Category and Across Categories.

- The Ultimate Collection of News Links (http://pppp.net/links/news/) is a reference source of media hyperlinks by region and country. Conduct searches of the appropriate news media source(s) in the country or area of interest.

- News Index (http://www.newsindex.com) is a search engine for current news stories from worldwide sources.

- Pandia Newsfinder (http://www.pandia.com/news/index.html) is a search engine for current news and links to news sources.

- AlertNet (www.alertnet.org) is a Reuters-sponsored service providing global news, communications and logistics services to the international disaster relief community. The public pages, accessible to anyone, carry a live news feed and articles describing how relief agencies are responding to the latest humanitarian crises.

- BBC World Service (http://www.bbc.co.uk/worldservice/).

- Economist Intelligence Unit (EIU, http://www.eiu.com) electronic data requires a subscription or fee for individual articles. However, the Data Services and ViewsWire web sites contain country specific historical and current political, economic and security information.
-
  - EIU ViewsWire (http://www.viewswire.com/) is a daily country analysis service covering 195 countries and providing an unrivalled and timely database of background information, analysis and forecasts.

Use standard "search strings" employing simple Search Engine Math or Boolean Logic commands such as those listed below.

- *name of the national intelligence service(s)*
- *country name* **+intelligence**
- *country name* **+counterintelligence**
- *country name* **+espionage**
- *country name* **+security**
- *country name* **+spy**
- *country name* **+surveillance**
- *country name* **+clandestine**
- *country name* **+secret**
- *country name* **+secret** +**police**
- *country name* **+underground**
- *country name* **+covert**

c. Criminal Threat Research *(15 Feb 02)*

BACKGROUND

Most often, research on criminal threats and the availability of drugs will be incorporated into a Force Protection (FP) Threat Summary (THREATSUM) or Briefing. Both of these products should clearly and concisely define the actual and/or potential dangers to a specific allied/coalition force, in a specific geographic area of operations during a specific period of time. ***In all cases, the relevant "raw" research documents should be saved to an appropriate file folder under the country or area of concern.***

GUIDELINES.

Identify the criminal threats and define the availability of drugs. Traveler warnings and annual crime/drug reports issued by the State Departments or Foreign Ministries of the western governments are probably the best source of general information about the significant risks to Allied/Coalition Forces.

- U.S. Department of State Travel Warnings (http://travel.state.gov/travel_ warnings. html).

- U.S. Department of State, Crime Control Fact Sheets (http://www.state.gov/ g/inl/crm/fs).

- U.S. National Security Council, International Crime Threat Assessment (http://clinton4.nara.gov/WH/EOP/NSC/html/documents/pub45270/pub45270 chap3.html).

- Canadian Department of Foreign Affairs, Travel Advisories (http://voyage. dfait-maeci.gc.ca/destinations/menue.htm).

- British Foreign and Commonwealth Office travel advice notices (http:// 193.114. 50.10 /travel/countryadvice.asp).

- U.S. Department of State, International Narcotics Control Strategy Annual Report (http://www.state.gov/g/inl/rls/nrcrpt/).

- United Nations, Global Illicit Drug Trends (http://www.odccp.org/global _illicit_drug_trends.html)

Define current criminal activity in the country or area of interest. The below listed databases, metasearch engines and news search engines/sites should be searched employing the search strings indicated in paragraph C.2.3.

- Vivisimo (http://www.vivisimo.com/) a high quality metasearch engine.

- Ixquick (http://www.ixquick.com/) a high quality metasearch engine supplementing Vivisimo.

- NewsTrawler (http://www.newstrawler.com/nt/nt_home.html) is a metasearch engine for news archives and articles from worldwide sources. Search categories are By Country, Across Countries, Category and Across Categories.

- The Ultimate Collection of News Links (http://pppp.net/links/news/) is a reference source of media hyperlinks by region and country. Conduct searches of the appropriate news media source(s) in the country or area of interest.

- News Index (http://www.newsindex.com) is a search engine for current news stories from worldwide sources.

- Pandia Newsfinder (http://www.pandia.com/news/index.html) is a search engine for current news and links to news sources.

Use standard "search strings" employing simple Search Engine Math or Boolean Logic commands such as those listed below.

- *country name* **+crime**
- *country name* **+robbery**
- *country name* **+mafia**
- *country name* **+assault**
- *country name* **+trafficking**
- *country name* **+narcotics**
- *country name* **+ drugs**

d. *Medical Threat* (Rev 08 May 02)

BACKGROUND.

Most often, research on medical issues will be incorporated into a Force Protection (FP) Threat Summary (THREATSUM) or Briefing. Both of these products should clearly and concisely define the actual and/or potential dangers to a specific allied/coalition force, in a specific geographic area of operations during a specific period of time. In all cases, the relevant "raw" research documents should be saved to an appropriate file folder under the country or area of concern.

GUIDELINES.

1. Identify the types of disease and other significant environmental risks present in the area of interest. Traveler warnings issued by the State Departments or Foreign Ministries of the western governments are probably the best source of general information about the actual and/or potential medical dangers to Allied/Coalition Forces.

- U.S. Department of State Travel Warnings (http://travel.state.gov/travel_ warnings. html).

- Canadian Department of Foreign Affairs, Travel Advisories (http://voyage. dfait-maeci.gc.ca/destinations/menue.htm).

- British Foreign and Commonwealth Office travel advice notices (http:// 193.114. 50.10 /travel/countryadvice.asp).

2. Determine if there has been a recent (with the last year) outbreak of disease or a rise in the incidence of other significant environmental risks in the area of interest. The following databases, metasearch engines and news search engines/sites should be consulted.

- World Health Organization (http://www.who.int/home-page/)

- Center for Disease Control (http://www.cdc.gov/travel/reference.htm)

- International Society of Travel Medicine (http://www.istm.org/) is a global organization dedicated to healthy and safe travel medicine and international care.

- U.S. Department of State HIV/AIDS and Emerging Infectious Diseases (http://www.state.gov/g/oes/hlth/).

3. If no significant data is found or more information is required, conduct searches using news search engines.

e. Geo-Political, Military & Country Information Research *(Rev 04 Mar 02)*

POLITICAL LEADERS

- U.S Department of State, Chiefs of State and Cabinet Members of Foreign Governments (http://www.odci.gov/cia/publications/chiefs/index.html) provides a listing of government officers.

MILITARY INFORMATION

- American Federation of Scientists (FAS), Military Analysis Network (http://www.fas.org/man/index.html) provides links to military order of battle information. Additionally, Search FAS (http://www.fas.org/) can be used to research the organization's databases.

- Jane's Information Group (http://www.janes.com/) is a **subscription service** providing access to highly credible data on geo-political, military, intelligence and terrorism issues. However, **free access is authorized to selected items** via Jane's Search (access through the homepage), the Sentinel Risk Pointers (http://www.janes.com/regional_news/sentinel/risk_pointers.shtml) and International News Briefs (http://www.janes.com/regional_news/international/news_briefs/foreign_report_toc_2001-1.shtml).

- Center for Defense Information (http://www.cdi.org) - This site has weekly articles compiled or produced by academics which include information on world terrorist, weapons of mass destruction, military defense, etc.

COUNTRY BACKGROUND INFORMATION

- U.S Department of State, Background Notes (http://www.state.gov/countries/, http://www.state.gov/r/pa/bgn/) provide historical geo-political, economic and social information on most nations of the world.

- Library Of Congress, Country Studies Country Studies On-Line (http://lcweb2.loc.gov/frd/cs/) presents a description and analysis of the historical setting and the social, economic, political, and national security systems and institutions of countries throughout the world. Information is not copyrighted and thus is available for free and unrestricted use by researchers. As a courtesy, however, appropriate credit should be given to the series.

- U.S. Central Intelligence Agency (CIA), World Factbook (http://www.odci.gov/cia/publications/factbook/index.html) provides access to historical social,

economic, political, and national security information on countries throughout the world.

- American Federation of Scientists (FAS), Country Studies (http://www.fas.org/news/index.html) provides background information on national security, military, intelligence and medical information. Additionally, Search FAS (http://www.fas.org/) can be used to research the organization's databases.

- Lonely Planet Travel Guide (http://www.lonelyplanet.com/destinations/) provides historical cultural and social information on the countries of the world.

- Economist Intelligence Unit (http://www.eiu.com) is a **subscription service** providing access to highly credible data on regional and national geo-political, economic and security issues. However, **free access is authorized to selected items** via the Data Services and ViewsWire web sites which contain country specific historical and current political, economic and security information.

    - EIU ViewsWire (http://www.viewswire.com/) is a daily country analysis service covering 195 countries and providing an unrivalled and timely database of background information, analysis and forecasts.

f. Geo-Political & Military Information – Russian Annex (Rev 04 Mar 02)

RUSSIAN MILITARY

http://www.webcom.com/~amraam/shipind.html - A site containing information on Russian ships and submarines currently active and in which fleet they belong. This site can be used as an additional source, or to supplement classified information.

RUSSIAN NEWS

http://en.rian.ru/rian/index.cfm - Russia Information Agency - A daily news site like BBC. Information appears on this site for two days before it is put into an archive and must be paid for.

http://www.gazeta.ru/english - Gazeta. This site contains Russian information that is updated on a daily basis.

http://www.interfax.com/com?item=Rus - Interfax. This site contains economic news from Russia.

http://www.itar-tass.com/newsdir.htm - ITAR-TASS. This site only gives headlines. The details of articles must be purchased. Small photos are also available.

http://www.russianobserver.com - Articles on this site are updated on an irregular basis. Checking this site weekly is sufficient.

http://www.therussianissues.com - Articles on this site are updated on an irregular basis. Checking this site weekly is sufficient.

http://www.smi.ru - Russian version of The Russian Issues. Can be translated using a web translator such as Babel fish.

http://www.themoscowtimes.com - In English. This paper is updated on a daily basis.

http://www.cdi.org/russia - This site contains the Russia Weekly, which is a weekly report compiled from articles from various sources. The site is updated on Fridays, and should be checked weekly. It also contains a link to Moscow Times.

http://www.adresa.ru - Another Russian site. Must be translated.

http://www.vesti.ru - Russian paper updated on a daily basis. Must be translated.

g. Gazetter, Port, Geodessy and Map Research *(Rev 02 May 02)*

GAZETTEER (Database of geographic feature names)

- U.S. National Imagery and Mapping Agency, GEOnet Names Server (GNS) Gazetteer (http://gnpswww.nima.mil/geonames/GNS/index.jsp) provides access to worldwide (excluding the U.S. and Antarctica) geographic feature names. For names in the U.S. and Antarctica go to the US Geological Survey (USGS, http://geonames.usgs.gov/).

- Global Gazetteer (http://www.calle.com/world/) is a directory of 2880532 of the world's cities and towns, sorted by country and linked to a map for each town. Presentation Copyright 1998-2000 by Falling Rain Genomics, Inc.

PORT INFORMATION

- U.S. National Imagery and Mapping Agency, World Port Index Pub 150 (http://pollux.nss.nima.mil/pubs/pubs_j_wpi_sections.html). Documents are in PDF format and require Adobe Acrobat Reader.

- World Port Directory (www.maxpages.com/port) is a commercial internet webpage providing information on selected ports of the world and waterways frequented by cruise ships.

MAPPING AND GEODESY PRODUCTS

- U.S. National Imagery and Mapping Agency Geospatial Engine (http://geoengine.nima.mil) provides access to imagery of the Earth, maps and other geospatial information.

- U.S. National Imagery and Mapping Agency, Aeronautical Information Homepage (http://164.214.2.62/products/digitalaero/index.html) provides access to global aeronautical geospatial information and services to include the Digital Aeronautical Flight Information File (DAFIF®), Flight Information Publications, FLIP Planning Documents and Enroute Supplements, and Prototype Terminal Instrument Procedures and Airport Diagrams.
  - Documents are in PDF format and require Adobe Acrobat Reader.
  - Also accessible via the NIMA Map and Geospatial Data web page (http://164.214.2.59/geospatial/geospatial.html).

- Earth-Info Navigation Page (http://www.earth-info.org/JavaGlobeStart.html), provides access to a world of imagery, maps, and other information. Hyperlinks are available to the CIA World Factbook (http://www.odci.gov/

cia/publications/factbook/index.html), ImageLinks, ImageNet, DOI-10 Imagery DTED Select DOI-10 Imagery, DTED® 0 Terrain Data, VMap 0 Vector Map, VMap 1 Vector Map, ONC Raster Map, TPC Raster Map and National Geographic geodesy products.

- Also accessible via the NIMA Map and Geospatial Data web page (http://164.214.2.59/geospatial/geospatial.html).

- U.S. Central Intelligence Agency World Factbook (http://www.odci.gov/cia/publications/factbook/index.html) provides access to regional and country geopolitical maps.

- Lonely Planet Travel Guide (http://www.lonelyplanet.com/destinations/) provides access to country and city tourist maps.