OSS 21 May 2000, Rosslyn, Virginia, USA

Information Overload.

By Steve Edwards
New Scotland Yard

Having been given 15 minutes to present the UK law enforcement picture as far as open source work is concerned I did not want to bore you unnecessarily with the few things I will say later. Instead, here are a few thoughts about things in general, open source, information technology, Internet, whatever.... Take them with a pinch of salt - or not. I hope they will give you all something to think about.

State of the Disunion.

We live in a technological age. We live in the information age. We live in a shrinking world in one sense we live in an expanding world in another. We live in an accelerating world.

Knowledge is power, so we hear, but more important what does the lack of knowledge mean?

Our 'societies' depend on people making informed decisions and making those decisions in quicker and quicker time. Individuals with that onus need to have larger pools of information available to them and the ability to access, filter, collate, analyse, disseminate and action that information in a timely, intelligent and productive manner. Failure to do this in the business sense can mean, embarrassment, litigation, competitive loss, bankruptcy or any number of financial disasters. Failure in the other sectors can mean far worse. It can mean bombing an embassy we should have known was there. It can mean the end of a political era, which we did not predict. It can mean the invasion of a country we are supposed to be protecting. It can mean the death of our armed forces when we might have predicted a different outcome. It can mean cult or terrorist attacks using special techniques and materials when we might have identified and neutralised all three.

Our *intelligence* agencies are fast becoming *information* agencies. It used to be that the definition of intelligence had some sort of relevance to it. Today, we require so much intelligence that it tends to become devalued in someway by its sheer volume and scope. Not knowing where the next focus will be means that we throw our nets in wider and wider circles. We have almost reached a point where, with global availability of information through the Internet, that we assume all information to be intelligence purely because of its accessibility.

This has tremendous implications for anyone working in the intelligence arena. Advances in storage technology, search and retrieval technology and information and communications technology have meant that the concerns raised in the 80's and 90's over costs and feasibility are now all but done away with. The concerns now are not about where do we keep it all and how do we access it, but more, how do I make any sense of it and what do I do with it.

Working as an open source specialist for 5 years now I have come to realise that despite all of the excellent intelligence gathered by conventional means; eavesdropping, surveillance, informants or agents - one can now produce enough intelligence from open sources alone to keep intelligence officers busy on a full-time basis. In fact, one might now begin to worry that open source intelligence may get ignored simply because there is not the capacity to action it.

In addition to this, we are frequently presented with an overload of information, which we know the customer will not have the time of resources to analyse. More and more we find ourselves filtering and re-filtering information, just to make it manageable, all the time running the risk of filtering *out* that vital 'thing' that the customer actually needs.

Looking at the availability of information today it would seem logical to expect the emphasis to be moving away from retrieval towards analysis. However, it is now not uncommon to make a search on the Internet and be presented with half a million hits. Even using skilled Boolean logic and the best search and meta-search engines this can sometimes stay in the tens of thousands. Who has the time to check all of that? But what is to say that the vital information isn't in article 6,432?

Making sense of it all, we hope, will be the next big thing. The last I heard was that 200 million pages of the Web were indexed by the best search engines leaving 800 million untouched. One does not like to think what the figure actually is but what it will yet become is another thing entirely.

As an information professional I am often asked which is the best search engine, I used to say AltaVista or Yahoo or one of the name brands, because they seemed to bring back so much more information. With so many options available and with so many different criteria used by each engine, more and more people are switching to meta-searching where only the top ten hits, or so, from each search engine are displayed. Anyone who has ever sent a query to 'Ask Jeeves' and compared the different responses from the different search engines will know what that diversity can mean. For the future, one might predict that success will lie, not with those who can find the most information, but with those who can bring back the most *relevant* information. It is amazing to think that even with the most responsible search engines, that the user is still left to his or her own devices when formulating searching techniques. Buttons which designate 'and' or 'not' or even the ability to do synonym or proximity searching (within 5 words or so) are still not common place on the Internet.

And here are a few more concepts (ramblings) to toss into the technological debate.

Auto Intelligence.

The conspiracy theorists tell us that 'Echelon' style applications are already doing it. But what is, 'it'... and for 'whom' are they doing it?

The problem? There is a world of information and/or intelligence out there and everyone wants a piece of it. Open source intelligence has been sold and the customer wants all he or she can get because it can be cheap, it can be cutting edge and it can be there now. So how do we service everyone's needs?

I have no idea how many people may ever read this article but however many there are I can pretty much guarantee that they will all have different intelligence needs, foci, and resources. Using law enforcement as a case study, let's look at the needs of an intelligence officer working in the field of organised crime. Put that officer up against a roomful of other officers similarly employed and what do you have? Drugs, firearms, money laundering, fraud, prostitution, protection, gambling, murder... the list could go on. Now split them into smaller groups. Take the drugs guys and girls into a room and see what you have. Hard drugs, soft drugs, commodity driven, target driven. Hydroponics, harm reduction, heroin or hash, they will all have a different area of interest.

So here's the concept. And remember where you heard it first.

Take the Internet as the largest single pool of globally available free information we have. Add to that any pay-as-you-go services which one might be subscribing to through the web. Stick in a couple of commercial on-line services and maybe a few jukeboxes of CD-ROM information.

Lay over the top of that a piece of software, similar in nature to an Internet news profiler, with the ability to take words, terms and phrases and retrieve relevant material. Make them intelligent agents so they work even when you are not on-line and give them the ability to sort by date, relevance etc. and filter out the unwanted material. The software will not only submit it's searches

to search and meta search engines but will use news archives, news trackers and any other existing useful web-sites or tools. It will also take the given criteria and retrieve it automatically to a database (maybe a terabyte or two) but will also regularly examine the database itself for currency and relevance, archiving anything that should not be there. It will have lifted entire web-sites, if necessary, in order to make later research as covert as possible.

Suddenly one has a pretty powerful intelligence database covering the subject of one's choice which not only puts all the information in one place where it can be analysed off-line but also grabs that web data forever, available even when the pages vanish from the Internet.

Imagine tasking such a beast to research, for example, terrorism. Take the US State Department's annual report on Patterns of Global Terrorism and use their list of terrorist groups as a start point. Add a few more from places like Perdue University or wherever. Then add a few key words like explosion, kidnap, assassination, etc. The software will, like so many web types do, auto suggest some extra terms one might not have thought of and off we go. Every search engine tasked to work for you, every on-line news agency, FBIS, BBCMS, Academic sites, images, video, sound, news groups, usenet, Janes, Lexis-Nexis, Oxford Analytica, Kroll Associates... we could go on. The possibilities are almost endless.

The beauty is that the researcher becomes the programmer and the machine does all the work. The intelligence data is stored forever (how many off us will have wished we'd kept all the news items and usenet ramblings about the African Embassy bombings when the next one happens?) and it can be searched off line without worries about security, footprints, cookies etc. etc.

Yes, I'm sure that somewhere, for some reason, such a beast might exist. I'm sure that someone somewhere may be getting a fine service from some very sophisticated hardware and software. The thing is that you or I aren't. We I know I'm not! And it's not difficult, this is not rocket science. The technology is simple, and it's already here and in use. A few tweaks to similar systems or a join between this concept and that is all that's required. It's simplicity is it's beauty. Bespoke intelligence databases can be automatically created to suit all sorts of needs at all levels.

The Rest of the Web

I see that the very first workshop in this conference will be dealing with the issue of accessing the 80% of the Internet not covered by search engines. So the presentation could be that my next point is completely redundant, or it could just help make the point.

If it really is true that only a small fraction of the Internet is indexed or covered by even the best search engines (call them what you will), what are we doing about it? I thought, in my naiveté, that there would be some hi-tech scheme somewhere, looking to focus on the rest of the stuff. Apparently not. I hear that EROLS, one of the preferred ISP's for the DC Suburban area is not picked up by any of the main search engines and that is just one of hundreds world-wide. What are all those beltway bandits using as their homepage service? OK, I probably know the answer to that one, but you see what I mean. If any government agency wants a project which will enamour itself to the rest of the intelligence community and, if it goes public, make itself a whole bundle of cash, then look at the rest of the Internet, plonk a search engine over it and make it available. People will bite your arm off for it.

And finally, just in case I have any credibility left at all...

The Global Brain.

When we look at the huge technological advances made in our lifetimes - or even our children's lifetimes, and look forward exponentially, we are presented with the huge potential of what we will be able to achieve. Only a few years ago, my first computer was a 25 MHz 486 processor with 100 Mb of storage space. And it was top of the range. Nowadays when we buy any type of It

equipment we automatically assume it to be obsolescent by the time we leave the store. We are now buying 1 GHz processors, virtually unlimited storage, CD Re-writers, DVD, Modems, ISDN - and just when you think they couldn't get any more bits down the line, someone comes up with a way of getting 5 times that amount crammed in there somehow.

We are already seeing the development of spectacle lenses which can show computer screen images. We already have speech recognition and text to speech. The technology gets smaller and smaller and needs less and less power. How long before the chip moves from outside the skin to inside the skin, using the body's own electricity as power and changing text to speech into our auditory system. When global communications become wireless and the Internet can be beamed into the chip anywhere in the world, connecting the human to the world-wide pool of knowledge, what is next? Is it fantasy or is it just a matter of time. I can beam my thoughts to anyone in the room who has a Palm Pilot simply by typing them out and hitting a button. I can do it on a cell phone or pager. Think ahead five, ten, twenty years. Think back twenty years. Where were we technologically? There is more computing power in the average automobile than there was on the first lunar module.

I hope you all have a good conference. Don't forget EuFis later in the year, and remember my Open Source caveat:

'Use it as corroboration - *never* use it without corroboration.'

OSS 21 PRIMER Essential Elements of Information Joint Planning, Operations Other Than War and Open Source Intelligence - Link Page

Previous	OSS '00 Major Patrick L. George, CCA, A Belgian Situation Report,
Next	OSS '00 Steve Edwards, Scotland Yard, Information Overload (Slides),

Return to Electronic Index Page