24 MONTH FIXED RATE CD

***6259

Your on the Street Reporter

Uyless Black

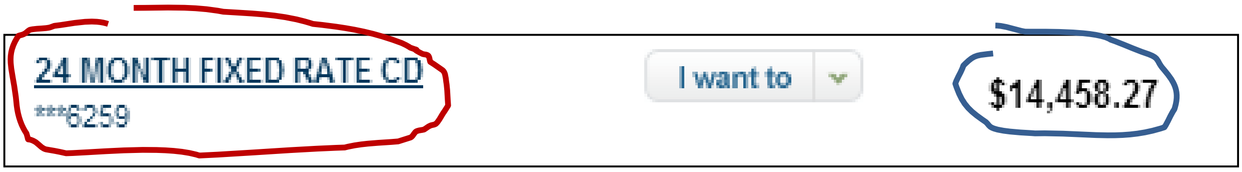
Metadata

Metadata

February 1, 2014

Hello from Your on the Street Reporter, continuing the series on Internet Privacy, NSA surveillance, and Google's googles. As the name of this piece states, the report deals with *metadata*: what it is, what it is not, and why you should know the difference.

I have come across several definitions and descriptions of the word metadata.[1] Using Figure 1, the general concept of metadata is easy to grasp. Metadata is information about other information. Stated another way, metadata is data about data. Metadata is not the data itself. In Figure 1, the information circled in red is metadata. The information circled in blue is data. I prefer the following description of this data inside the red circle: It is user content.



24 MONTH FIXED RATE CD
***6259

$14,458.27

Figure 1. Metadata and user content.

The information of "24 MONTH FIXED RATE CD ***6259" is data about data. It identifies that account number 6259 is a twenty-four month fixed rate certificate of deposit (CD). The data itself, again what I call user content, is the value of the CD: "$14,458.27."

You may be surprised about this distinction between metadata and data. If metadata is free game for NSA or Google to examine, then this metadata reveals a lot about the party who owns account 6259: He or she… or it, such as a corporation, a CD ID theft group, a terrorist group, or your ex-spouse. Whoever the party is, the party is investing in safe CDs. The party has a conservative bent, at least for this specific metadata record. The party is looking into the short range lens of life for this specific investment: two years. The party is reasonably well-established. After all, the party has sufficient credentials to open an account at a financial institution.

This one metadata record reveals only tidbits about the owner of this CD. However, if a snooper can capture *all* the records of this party, the snooper can glean and infer a considerable amount of intelligence from this so-called non-personal data. (A topic covered in a later report under the name *Big Data*.)

[1] I've been using the term since my days working in data-base management systems in the early 1970s. We programmers stored data on files and developed systems to explain, describe, and access the data. We called these systems *metadata*.

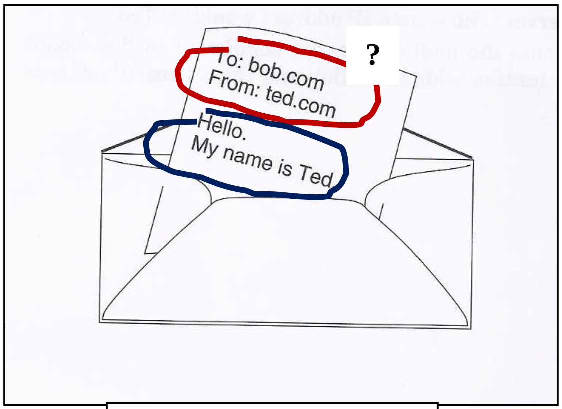


Figure 2. Not so sure.

For this article, in Figure 2, I have circled in red what could be called metadata, but with a question mark to note disagreement on this distinction. I have circled in blue what is *not* called metadata. The images inside the blue circle are called *user content*.

However, some experts state that *all* this information is user content, including the “To” and “From” information.[2]

They distinguish metadata from user content by: Anything the end user enters on his/her keyboard, pad, etc. that is placed inside an envelope (the Internet IP packet) is user content and should be subject to privacy protection. I agree with this distinction. There may be exceptions to this general approach (which I hope my readers will point out in your responses to these reports). If so, these exceptions can be specified in a court order.

I think you and I can say with confidence, this idea---that “bob.com” and “ted.com” are personal tidbits of information and not subject to search---is anathema to snoopers such as the NSA and Google. Reasonably so, as these addresses often reveal much information about the two parties. For example:

Metadata can reveal a lot about the traffic. By accessing the “To” and “From” information, NSA (assuming bob.com and ted.com are identifiable names of Bob and Ted) instantly possesses intelligence (maybe for later mining) about Bob and Ted: Namely, they communicate with each other.

For the postal service, the content on the envelopes is metadata. The “To” and “From” addresses on the envelope are metadata. Everything inside the envelope is user content. But as seen in Figure 2, this distinction does not hold for Internet traffic. Regardless of the opinion you or I may have on this subject, the “To” and “From” email “addresses” are considered metadata and open for examination.

The important point to repeat for this discussion: *Metadata is not considered to be user content. Therefore, it can be used by NSA for surveillance and Google for marketing*. My AOL address of ublack7510@aol.com is open for anyone to see and use. For my AOL address, it has information about me. (However, I could have an address that does not reveal anything about me, such as 7510@aol.com. This address is actually the address of a node [such as a computer] that I am using. Regardless of what it identifies, it is still open for inspection and use.)

[2] Thank you, Paul, for bringing up this point of distinguishing between metadata and user data.

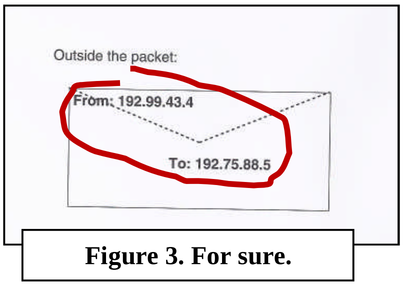
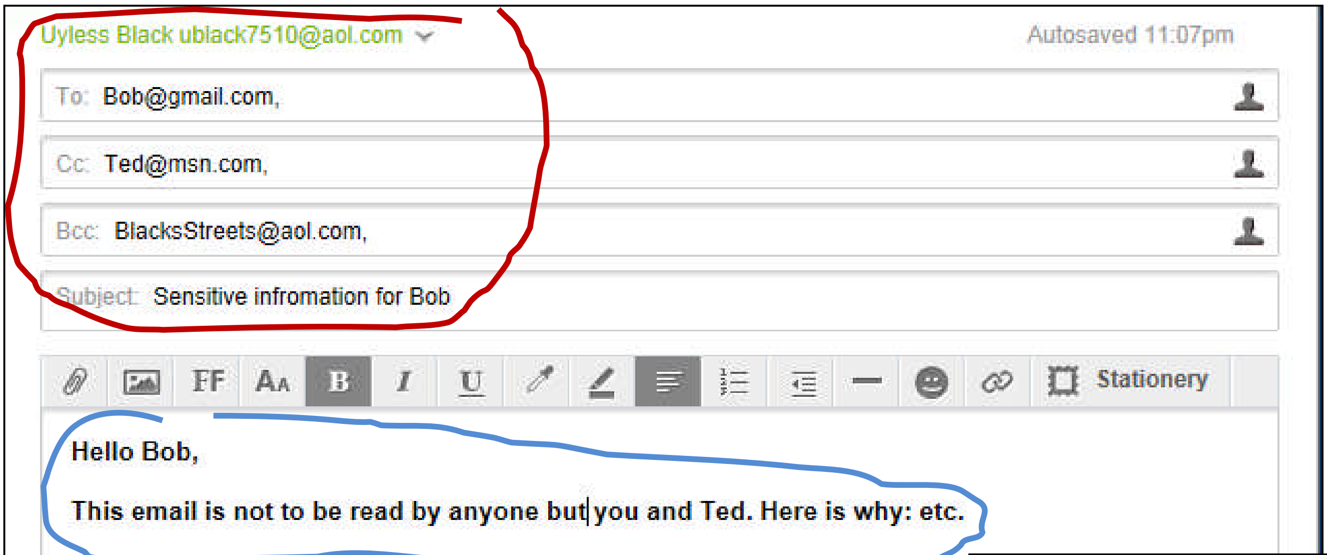


Figure 3. For sure.

Let's return to another figure from the last report, as seen in Figure 2. We learned another identifier, the IP address, is used to route our email and other traffic to and from machines, such as our computers. Keep in mind that inside this packet is the email shown in Figure 3 (and for that matter, Figure 1). The red circle indicates that IP addresses are considered to be metadata.

The question is simple: Do you care if the email identifiers of bob.com, ted.com and the IP addresses are not private? Do your care if they are treated as metadata? For IP addresses, forget about it, as you have no say in the matter (nor should you), as they are managed by the Internet system (like ZIP codes in the postal service). How about Bob.com, or in my situation, ublack7510@aol.com?

Before answering, consider the email in Figure 4. The only information in this email that is not viewed as metadata is "Hello, Bob, This email is not to be read by anyone but you and Ted. Here is why: etc." The email addresses are not private, no surprise there. But what might surprise you is that the Subject line is also considered metadata. It can be read and used by anyone who is reasonably knowledgeable about monitoring Internet traffic. Obvious examples are the NSA and Google. Less obvious examples are parties that conduct commercial espionage, blackmailers, credit information thieves, and other never-do-wells.


Uyless Black ublack7510@aol.com Autosaved 11:07pm

To: Bob@gmail.com,

Cc: Ted@msn.com,

Bcc: BlacksStreets@aol.com,

Subject: Sensitive infromation for Bob

Hello Bob,

This email is not to be read by anyone but you and Ted. Here is why: etc.

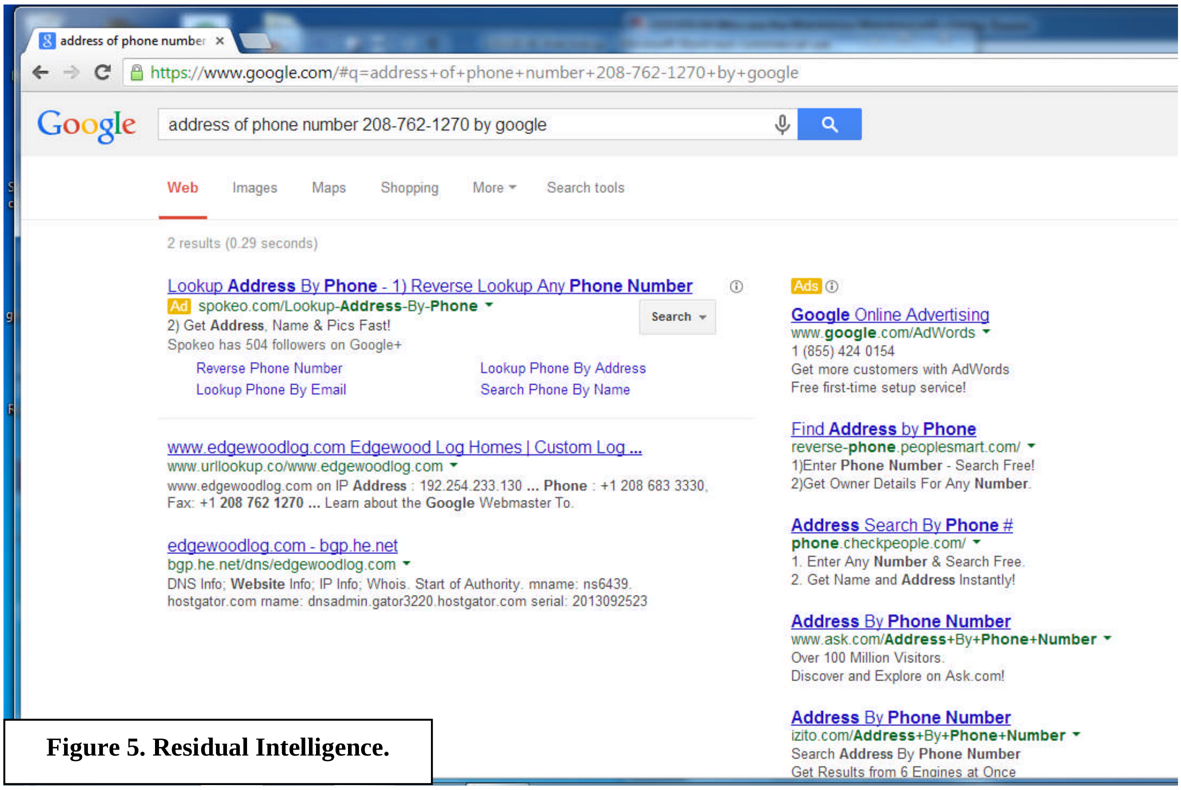
Figure 4. Red: Metadata.

Should you care? That is a question only you can answer. Just keep in mind that a lot of information about your user content can be gleaned from the Subject line and those email addresses. As well, for organizations, such as the NSA and Google, storing and examining perhaps all your emails can reveal a great deal of information about the activities in your life, indeed, about your personal life. Accumulated data, even metadata, can yield a lot of intelligence about someone or something. And bear in mind that these organizations can store your user content for possible later use. That is the rationale the NSA has offered by capturing and storing both metadata and user content.

Of course, you can place nothing or nonsense in the Subject line. But your friend receiving the email might not appreciate your cryptic nature.

More than Meets the Eye

To help gain an understanding of how much personal information can be gained by looking only at metadata, consider Figure 5. I keyed into the Google search engine my phone number (a land line number): "address of phone number 208-762-1270 by google." I placed in the request "by google" in an attempt to retrieve only Google's material. And look what I got back: several links that allow me to obtain a considerable amount of personal information about a piece of metadata: a mere phone number.


address of phone number ×
https://www.google.com/#q=address+of+phone+number+208-762-1270+by+google
Google
Web
Images
Maps
Shopping
More
Search tools

address of phone number 208-762-1270 by google

2 results (0.29 seconds)

Lookup Address By Phone - 1) Reverse Lookup Any Phone Number
Ad spokeo.com/Lookup-Address-By-Phone
2) Get Address, Name & Pics Fast!
Spokeo has 504 followers on Google+
Search
Reverse Phone Number
Lookup Phone By Email
Lookup Phone By Address
Search Phone By Name

www.edgewoodlog.com Edgewood Log Homes | Custom Log ...
www.urllookup.co/www.edgewoodlog.com
www.edgewoodlog.com on IP Address : 192.254.233.130 ... Phone : +1 208 683 3330, Fax: +1 208 762 1270 ... Learn about the Google Webmaster To.

edgewoodlog.com - bgp.he.net
bgp.he.net/dns/edgewoodlog.com
DNS Info; Website Info; IP Info; Whois. Start of Authority. mname: ns6439. hostgator.com rname: dnsadmin.gator3220.hostgator.com serial: 2013092523

Ads

Google Online Advertising
www.google.com/AdWords
1 (855) 424 0154
Get more customers with AdWords
Free first-time setup service!

Find Address by Phone
reverse-phone.peoplesmart.com/
1)Enter Phone Number - Search Free!
2)Get Owner Details For Any Number.

Address Search By Phone #
phone.checkpeople.com/
1. Enter Any Number & Search Free.
2. Get Name and Address Instantly!

Address By Phone Number
www.ask.com/Address+By+Phone+Number
Over 100 Million Visitors.
Discover and Explore on Ask.com!

Address By Phone Number
izito.com/Address+By+Phone+Number
Search Address By Phone Number
Get Results from 6 Engines at Once

Figure 5. Residual Intelligence.

Based on obtaining this screen, I did additional queries. The searches came back with my accounts on the social networks, on places of former residence, on my job history, etc. The amount of information I gained about myself was partly a function of how long I wanted to stay at it and do more surfing.

Should you care? Again, that's your call. For myself, I am resigned to the fact that the Internet will never have the privacy of the postal service. The toothpaste is out of the tube, and the Googles and NSAs of the world will make sure it stays that way. *I can live with this degree of intrusion, if the user content, the data inside that blue circle in Figure 4, cannot be examined unless a warrant is issued. We live in a dangerous world. Like it or not, we need NSAs.*

There is small chance of that happening now. Ironically, as of this writing, in America, anyone but Uncle Sam can read, store, filter, examine, and cross-correlate our user content. Google routinely does it for marketing purposes.[3] The NSA would like to do it to catch the bad guys. I heard one government official say, "It isn't the NSA a citizen should be concerned about. This agency is constrained by law to respect the Fourth Amendment. Google is not." Here are some thoughts from Google.[4]

> "Google also scans the content of emails of users of its Gmail webmail service, in order to create targeted advertising based on what people are talking about in their personal email correspondences.
>
> Shailesh Rao, country head of Google India, explained how Google scans the text of Gmail messages in order to filter spam and detect viruses. And while Google does this, the filtering system also scans for keywords in users' emails which are then used to match and serve ads right when you are reading your mail. So, when the user opens an email message, Google servers instantaneously display germane information that is matched to the text of the message."

In a later report, I will explain how you can use encryption to guard your user content. By using software we can gain back our Fourth Amendment rights, at least against petty hackers. Uncle Sam should have enacted legislation to protect our electronic letters, but did not. We have to do it ourselves.

[3] And the U.S. government can do it with a court order.

[4] Priyanki Joshi, "Every Move You Make, Google will be Watching You," *Business Standard*, March 21, 2009.