All.Net Analyst Report and Newsletter

Welcome to our Analyst Report and Newsletter

Provenance

I have been working on digital forensics in a number of legal matters lately and with the provenance research team in the InterPARES Trust research program in archives and records management (ARM) dealing with trust in online records and funded by the Canadian government at the University of British Columbia. It seems like provenance of digital records is problematic across the board.

The three questions

In looking at this combination of things, I have come up with my simplistic approach to dealing with the issue of provenance. From the perspective of any individual or institution, the key to understanding this issue, to me, seems to be in answering three questions:

1. How did it come to be?
2. How did it come to me?
3. What did I do with it?

If these questions are answered every step of the way, from conception through final delivery, and if the documentation is adequate to the need, and if the system of records is documented based on governance and tracked through execution, we may then use consistency analysis to evaluate the provenance in context and make a reasonable determination about the reliability of the record and thus perform a sound job of diplomatic analysis.

There lies the rub

However, history suggests that these set of conditions are rarely met. While professionals in ARM are generally aware of these issues and tend to address them to a greater extent in public and similar juridical records, other records are not generally kept to the same standard. In particular, records of most corporate entities I have examined tend to be sloppy in the extreme. In many cases these records are introduced in legal matters through the business records exception associating reliability with records kept and used in the normal course of business.

In addressing provenance, I generally try to get the answers to these three questions, and I very often find that the answers are either, in effect, unknown or undisclosed. When I say undisclosed, I don't mean to say withheld per se. Rather, that as a society, we generally accept what is given to us from 3 rd party providers who don't disclose the answers to these three questions in their normal course of business, and we pass this along to others with the same level of care and attention.

I don't think it's about cost

I have heard those in the ARM community tell me that the cost of keeping details such as these is a limitation on such systems. Indeed most ARM institutions I have been exposed to have woefully little resources for the relative criticality of their function to society or their institution. But I also think cost is used as an excuse for laziness in process design.

Copyright(c) Management Analytics, 2016 - All Rights Reserved

I don't mean to imply that the ARM profession is lazy. Quite the opposite. The problem in my view lies in the process and the difficulty of tracking high volumes of minutia related to high volumes of data in a manual process. While we have the basic notion of automated data processing built into our history, the design of modern information technology and the systems that stem from it are not often built with the intent or even awareness of provenance or use of records.

People cannot do this job, but computers can

There seems to be a push toward building “intelligent” computers. But computers are typically pretty bad at being intelligent, while people are really bad at mindless repetition. History suggests that people in basic data entry make about 20% erroneous entries when no controls are used to detect and correct such errors. Really bad at repeating the same task again and again. Computers, as bad as they are at dealing with such inputs, commonly make far less than one miscalculation per hour while performing billions of calculations per second.

When it comes to provenance of records in computers, there is really no reason people cannot design such computers to include the answers to questions 2 and 3 for almost any records they receive. For records they generate, question 1 should also be easy to answer.

So why don't computers do this for us?

People program computers. And the people doing these tasks are really bad at doing the things that produce provenance records. In fact, people are generally unwilling to document what they do to the point of writing thousands of lines of computer program code without a single comment or documented purpose or function. As a programmer at times, I plea guilty to the same crime. I figure it's obvious from the code. And it is, as long as you know what it does and how it works. I keep telling myself that as I explain my code to others for the eighth time.

Somehow, we cannot bring ourselves to do the basic maintenance tasks required to assure we know what we have and had and what we did with it. This extends to all areas. People are notoriously bad at doing backups. And even automated backups are still not done as well as what would seem to make sense. We have been doing computer backups since at least the 1950s – that's 60 years or so ago. And yet we don't really have that problem solved. We have a 90% solution. But 90% of the time for something that happens billions to trillions of times per day means billions to trillions of failures per day.

Maybe you could drive my car

The first fully automated self-driving car fatality happened in the last few weeks. I don't know the details yet, nor does anyone else. But regardless of the details, we do know that lots of people are killed every year in car accidents using human drivers. And perfection is not what we are ever likely to get, even if we always seek to achieve it. But driving is an incredibly complicated thing to automate for a car on the open road today by computer. Far more complicated than doing reliable backups. And yet we still cannot do reliable backups, especially with provenance.

I have a suggestion. Suppose we solve the really quite important 70-year old problem of backups really well just to convince ourselves that we are ready to try for automated driverless cars? The answer to this question? Likely that it is too boring to get to the last little bit of solving the backup problem. Programmer want something more exciting...

Copyright(c) Management Analytics, 2016 - All Rights Reserved

Surveillance and provenance

Provenance can be largely solved by pervasive surveillance. And this is getting more and more feasible today. It's not just the CCTV cameras all over the place. It's all the computer surveillance. And increasingly, we are seeing hardware architectural solutions that avoid all but the most sophisticated threats in surveilling and recording essentially all inputs to and perhaps outputs from computers. These are chips that go on the computer boards and record all or selected portions of the bus activity used to communicate between devices and the CPUs. Every keystroke, every sound that appears at the microphone, every bit stored or retrieved, all timestamped and recorded for offload to a remote facility whenever connectivity is available, all the time, not bypassable except by physically disabling the hardware device on the computer. And likely this will cause the computer to stop working.

Put this in every computer everywhere all the time, add in all the other surveillance, and pretty soon you start to have a pretty darned good history of human life at the resolution of the current sensor technology. Storage, as it turns out, is not the problem you might think it is. Between the massive available storage and the compression technology, it's pretty easy to see that enormous amounts of data could be collected, stored, and made available on demand.

Social implications

There lies the rub. If we turn our society into a surveillance machine, we will likely have a record of most everything relevant to establish provenance in almost all cases. Of course the effort of doing so will rarely be worth the cost. But we will have provenance.

In most places, this gets a big “so what”. The implications of living in total surveillance all the time seem much greater to me than the implications of having better or worse provenance. Of course power and money will still rule and prevent exercise of provenance against the rich and powerful while making it a club against the rest of us.

But if you look at the elections and unrest around the World in combination with the information now available that far exceeds anything in the history of humanity, one thing you will see is the prevalence of deception and propaganda. And if you look closely, you will see that the presence of facts seems to have very little effect. Do you believe what you saw with your own eyes, or what I tell you? Sadly, the answer is very often “what I tell you”. “Don't confuse me with the facts” seems to be pretty much the norm.

The exception – for now – is the legal system

For all of its imperfections, the legal system is, as far as I can tell, the only place that really takes the time to evaluate facts and separate out the rumors, lies, and other ridiculous stuff. Not that it is perfect or even close to it. But admissibility of evidence at least has a chance of separating the wheat from the chaff. More probative than prejudicial. Reliable and accurate. Demonstrably so under close scrutiny by experts. Presented by witnesses with credibility and expertise or personal knowledge of the facts. Everything else doesn't count.

Conclusions

I am interested in provenance. In particular in how deceptively simple it is. How did it come to be? How did it get to me? What did I do with it? Answer those 3 questions definitively and you are there. But what constitutes an adequately certain answer and how to get it, we don't know.

Copyright(c) Management Analytics, 2016 - All Rights Reserved