

# Taking Work Home

- *The Stolen Laptop, And Other Travails*

- *Authoring Style Using RUST Technology*

## Table of Contents

Taking Work Home.....	1
- The Stolen Laptop, And Other Travails.....	1
- Authoring Style Using RUST Technology.....	1
Introduction.....	1
Methods.....	3
Results.....	4
Discussion.....	8
Meta Data.....	10

## Introduction

### Security in Storage and Transport of Web Documents

Your laptop has just been stolen; should you panic ? The answer, unfortunately, is : HELL YEAH! And then of course panic subsides and you begin the painstaking process of rebuilding your new laptop into something worth stealing (again).

Slow down. Someone, if not you then probably your employer, still worries about the old laptop.

A stolen laptop is only the "end of the world" if the only world the data knows is the laptop. If you backup data in external (network) storage you have solved the rebuilding problem. This is something you should do. However, the only problem solved by backup is the ease of the rebuilding process, that someone is still worried about the old laptop, now in unfriendly hands.

### Computer Memory Portability and Work-In-Progress

A body of data retains value when it is stolen. When a document is transported over a network, encryption techniques are very effective. Unfortunately the reason encryption is effective do not apply to the case of the body of data contained on a stolen laptop or even a thumb drive. Both the volume of data encrypted with the same algorithm and the time

available to study a given algorithm play a large role in the success or failure to crack a code or password and, unfortunately, laptops contain lots of data and thieves often have lots of time. Every SGML type markup language (HTML, XML etc.) makes a common, possibly encryption beating "mistake". Element tags are repeated, a transmission habit quickly exploited. In addition, Operating System or Application configuration files with known plain text content but co-resident and encrypted can provide clues to decryption techniques. Theoretical measurements of encryption strength based only on complexity often do not tell the whole story.

### **Bimodal Markup - The RUST Tag**

The RUST tag is an XHTML tag, and needs an XML document framework for full functionality. The RUST tag is bimodal, that is it can produce two different markup streams (HTML documents), a source copy or a sanitized version, from the same source. Either data stream can be encrypted for transport across a network, but the value of the body of data in the two streams greatly differs. RUST is an acronym for **R**edact **U**nless **S**tatic **T**ext.

### **Telecommuting: Field Trips and Fishing Trips**

Various authors, journalists, and salesmen, who shall remain anonymous, believe that the biggest advantage e-mail has over voice communications is the lack of background noise; particularly the sounds of laughter and clinking glasses. Likewise, if trout talked, Bluetooth® technology would be marginal at best, but don't tell your boss that, instead tell her that the laptop she bought you only contains the details you really need to labor 24 X 7. You can also tell her that you would never participate in a conference call in PJ's and mismatched bunny slippers, the difference being that the former claim might be truthful.

## **Methods**

- I. Secure Archive (Backup Storage)
- II. Accurate Reference
- III. Tag Syntax

## Results

### I. Secure Archive (Backup Storage)

Diaries, Web Logs (and other logs), Reports etc. should be accurate in detail, it goes without saying, but need not be interlinked, referenced in detail at all times. "Treason is a matter of dates" implies both temporal accuracy and historical perspective.

Source documentation should be uploaded to a secure archive as soon as practical but can be replaced by the sanitized version on a local machine. In combination with an author's memory of events, the sanitized version retains much of its referential potential but substantially less intrinsic data value. Should the local machine be compromised or stolen, decryption will be no less possible, but much less fruitful.

The same would be the case, for example, where a researcher concerned only with tabular statistics "rusted" a column of Names, Social Security Numbers, or other ancillary non-correlative data from their local copy of raw data.

For example:

RUST Tags Unprocessed (Office Server)			RUST Tags Processed (Local Laptop)		
SSN	Age	Gender	SSN	Age	Gender
123-45-6789	67	M	###-##-####	67	M
234-56-7890	42	F	###-##-####	42	F
345-67-8901	53	M	###-##-####	53	M
...	...	...	...	...	...

The table, a calculation matrix, still indicates the average age, and gender proportion, etc., but the identification information is absent from the laptop copy. Note that *table rows* cannot be "rusted" in the same way.

### II. Accurate Reference

The concept of an "accurate reference" is quite a bit fuzzier than we like to think. A reporter's notes need only include the W's (Who, What, Where, When, Why/How) irresolvable at a later date. Put another way, we know both what we know and what we know where to look up.

For Example:

- (Published) "The deceased, Mr. -----, who declined to provide his mother's maiden name to Credit Bureau Authorities, died at home Tuesday. 'Quite shamelessly unappreciative of our needs'; whined a Credit Bureau spokesperson".
- (Notes) "Mr. Smith-Jones died at home Tuesday from an e-Bola virus he contracted from a discount PDA. 'Quite shamelessly unappreciative of our needs'; said Buck Howe of Dewey, Squeezem and Howe LLP."

But seriously, disclosure of information regarding a minor or disclosure of a cause of death from a genetically inherited ailment can have consequences for the living. Euphemisms and "Spin" have their legitimate uses when the class of (undisclosed) meta data remains unchanged.

The difference between "News" and History is resolution of the references, in history.

\_\_\_\_\_ discovered the \_\_\_\_\_ while sailing the Caribbean.

*Columbus discovered the Indies while sailing the Caribbean*, but that would not have been anybody's first guess at the time; the point being it was unnecessary for Columbus to correct the "mistake" himself, in his own time; nor even exactly when in time the spacial "mistake" was corrected. Data lost to a local machine is not lost to History which is the only loss that really matters. This is a general feature of *meta data*, although the data might be redacted the blanked space has only a limited class of values which fit. The class of the redacted data can be often be given to retain semantic meaning while still affecting a redaction.

[bystander] discovered the [location] while sailing the Caribbean.

But, *Superman discovered the Kryptonite while sailing the Caribbean*, well, that's just plain wrong. Everybody knows Kryptonite is a mineral, not a location and no comic book with an ounce of Journalistic Integrity would dare such a crude abuse. Seriously. In fact, analysis of these lexical constructs is possible to a limited extent (see: [Anchor•Buoy•Boat](#) Diagrams)

### III.Tag Syntax

Mark-Up Language : (X)HTML

Source Word Processor : Simple Text Processor (Notepad.exe etc.)

Processing : XSLT (Rust.xsl)

#### Inline (note:mixed content not allowed, text only)

1. `<rust method="mark" mark="#">Item</rust> » "####"`
2. `<p><rust method="tcell" tcell="Selection">Item</rust></p> » <p>Selection</p>`

#### Table Data (note:mixed content not allowed, text only)

3.
  - a. `<td><p><rust method="mark" mark="#">Item</rust></p></td> » <td><p>####</p></td>`
  - b. `<td><rust method="tcell" tcell="Selection">Item</rust></td> » <td>Selection</td>`

Mark-Up Language : Rich Text "What You See Is What You Get" [WYSIWYG]

Source Word Processor : OpenOffice Text (\*.odt)

Processing : None (Print or Export Selected Text)

n.b. The text is heavily stylized. A placeholder (bibliographic mark) is used to represent the text which will be used in a copy of the source or a reformat. The placeholder must nonetheless obey the RUST Principal, that is, *any text marked for redaction must only transfer the bibliographic mark to copies*. As can be seen above, the `<rust>` tag obeys the RUST Principle because the *mark* attribute only is transferred while the original content is

redacted.

In the OpenOffice OpenDocumentText display, a bibliographic mark default looks like this : **Bibliographic entry**, although different styles can be used to denote the purpose or type of the mark. For example:

- Citation – **Bibliographic entry**
- Microformat Tag (meta data) – **Bibliographic entry**
- Personal Identifying Information (PII) – **Bibliographic entry**

This style will have no effect upon the style of the mark in a copy, but will make the source easier to read.

As a \*.odt file, the tabular example above would look like this :

RUST Tags Unprocessed (Office Server)			RUST Tags Processed (Local Laptop)		
SSN	Age	Gender	SSN	Age	Gender
<b>Bibliographic entry</b>	67	M	[alpha]	67	M
<b>Bibliographic entry</b>	42	F	[alpha]	42	F
<b>Bibliographic entry</b>	53	M	[alpha]	53	M
...	...	...	...	...	...

The table still indicates the average age, and gender proportion, etc., but the identification information is absent from the laptop copy. Note that *table rows* cannot be "rusted" in the same way.

Note that a cell of a spreadsheet may not contain a **Bibliographic entry**. This is, as they say, a feature, not a bug. When you copy a table from a source \*.odt file to a spreadsheet \*.ods file via the clipboard, only the mark is transferred. Other behavior would violate the RUST Principle.

In OpenDocumentFormat (ODF 1.0) XML the tagged text is (content.xml) :

```
<text:bibliography-mark
  xmlns:text="urn:oasis:names:tc:opendocument:xmlns:text:1.0"
  text:identifier=""
  text:bibliography-type="class-name"
  text:report-type=""
  text:howpublished=""
  text:title=""
  text:address=""
  text:annotate=""
  text:author=""
  text:booktitle=""
  text:chapter=""
  text:custom1=""
  text:custom2=""
  text:custom3=""
  text:custom4=""
  text:custom5=""
  text:edition=""
  text:editor=""
  text:institution=""
  text:isbn=""
  text:journal=""
  text:month=""
  text:note=""
```

```
text:number=""
text:organizations=""
text:pages=""
text:publisher=""
text:school=""
text:series=""
text:url=""
text:volume=""
text:year="">[class-name]</text:bibliography-mark>
```

Note that in the <rust> tag content is content and the mark is an attribute whereas in the <text:bibliography-mark> tag the mark is content and the attributes specify the content(s) of a data base table column. Both representations are equivalent and both are optimized for the editing context – one can always view HTML as text but ODF XML cannot be viewed directly from the Word Processor.

The HTML <cite> tag is constructed just like the ODF <text:bibliography-mark> tag. In HTML Readers (web browsers) the tag does not obey the RUST Principle because the attribute "title" containing citation data is not redacted in the plain text source and is often rendered as a "tool tip". The same is true of the HTML abbreviation element <abbr> and HTML acronym element <acronym>. For all three elements, one could use the "id" attribute to store referenced content however an element's "id" is unique in a valid HTML document. Several available attributes might seem suitable at first, but all ultimately fail. The reason for failure is not element syntax but rather the nature of relationship links in HTML and the model used for mapping data. A Resource Definition Framework (RDF) is only possible in the absence of a "reverse" link from class to instance (meta) data – one can put many books in a library but a library is not just a pile of many books. The difference between subjectivity and objectivity can not be computed although it can be reconciled. The usual way to define fixed references is to collect the different possible types into a [definition space](#), or a [name space](#).

## Discussion

The RUST tag can be used to enforce security policies dealing with data and information held on local machines and to break links to a historical archive thereby protecting an internal data reference from full disclosure.

There are three cases for the substitution of Text.

- i. A single character substituted for a multi character String.
- ii. A multi character String substituted for one or more characters.
- iii. A single character or String substituted for a Table Cell.

It can be argued on symmetry grounds that there must be an even number of transformation cases, and indeed there is a fourth case — a *null* substituted for another *null*. Meta Data and metaphysics sometimes intermingle. What the “even case” or *trivial case* implies is that data cannot be re-manufactured with certainty once lost. For the particular goal of redaction, this case is trivial in name only as the lack of an inverse transform prohibits replication of the original document.

There is no way to use the RUST Tag with formal meta data, that which is segregated from the body of the document or invisible on display. This is intentional. The purpose of the RUST Tag is to protect personally identifiable information and other sensitive meta data, not to keep documents secret.

Table Cell data is problematic for two reasons.

- i. Tabular data often represents a calculation matrix from which statistical measures of the data are derived. Only identifying information, in a column, is suitable for redaction.
- ii. If a cell column is to be “rusted”, then the content of the cell must be text only. Only atomic data is suitable for redaction.

Web Page authors have been warned for years not to use tables for document layout. Paper is two dimensional and anything printed is a 1 x n row table of sections. Now, a table within a document occupies one of the n rows — not a problem. A document in a table, on the other hand, is a problem because every document contains at least two rows (Meta Data + Data or HEAD + BODY, etc.), and there is no way to pivot this for printing. Seen in this way computer memory (instructions + data) fully qualifies as a “document” as does formatted audio or formatted video — but not, for example, white noise, an n x n design or silence a 1 x *null* design.

Just in case the previous paragraph causes a reader to get dollar signs in their eyes, the author has some doubts that the "SONY Challenge" is even possible. If an audio engineer were to introduce noise at some frequency that identified a song the result would fit the 1 x n design, but unfortunately the identification noise layer would still be *redactable*. Admire people who can *Name That Tune* in one note, but don't pay them royalties unless you have too much money. Way too much money.

Meta data stored in the (non-displayed) front matter or header is still subject to propagation and proliferation when the file format is changed, but nearly always ignored for printing on paper. The RUST Technology takes a different approach by limiting propagation in most file format changes as well as print on paper of all meta data, displayed or otherwise. This approach is particularly appropriate to *Personally Identifiable Information (PII)*.

## Meta Data

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:dctype="http://purl.org/dc/dcmitype/"
  xmlns:pii="http://purl.org/pii/terms/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcam="http://purl.org/dc/dcam/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <!--DCMI Abstract Model-->
  <!--Instance Document-->
  <rdf:Description rdf:about="http://www.RUSTPrivacy.org/TakingWorkHome.pdf">
  <dcterms:title xml:lang="en-US">Taking Work Home - The Stolen Laptop and Other Travails</dcterms:title>
  <dcterms:identifier>
  <rdf:Description>
  <rdf:value rdf:datatype="http://purl.org/dc/terms/URI">http://www.RUSTPrivacy.org/TakingWorkHome.pdf</rdf:value>
  </rdf:Description>
  </dcterms:identifier>
  <dcterms:language>
  <rdf:Description>
  <rdf:value rdf:datatype="http://purl.org/dc/terms/RFC4646">en-US</rdf:value>
  </rdf:Description>
  </dcterms:language>
  <dcterms:source>OpenOffice.org/2.3$Win32 OpenOffice.org_project/680m9$Build-9238</dcterms:source>
  <dcterms:creator>RUSTPrivacy.org</dcterms:creator>
  <dcterms:issued>
  <rdf:Description>
  <rdf:value rdf:datatype="http://purl.org/dc/terms/W3CDTF">2007-12-01T17:04:21</rdf:value>
  </rdf:Description>
  </dcterms:issued>
  <dcterms:contributor>Gannon J. Dick</dcterms:contributor>
  <dcterms:modified>
  <rdf:Description>
  <rdf:value rdf:datatype="http://purl.org/dc/terms/W3CDTF">2008-05-17T11:33:56</rdf:value>
  </rdf:Description>
  </dcterms:modified>
  <dcterms:provenance xml:lang="en-US">
  Printed By Gannon J. Dick on 2008-05-17 in en-US.
  </dcterms:provenance>
  <dcterms:subject xml:lang="en-US">Authoring Style Using Redact Unless Static Text (RUST) Technology;Personally Identifiable Information;Meta Data;Mark-
  up;XHTML;PII;Security</dcterms:subject>
  <dcterms:description xml:lang="en-US">Word Processors include mechanisms to insert content, delete content and, for collaborative work, versioning schemes.

  To protect personal privacy and protect legal rights however, the author believes it makes more sense to embed a "redaction command" in the content of a
  document.</dcterms:description>
  <dcterms:publisher>mailto:gannon_dick@yahoo.com</dcterms:publisher>
  <dcterms:rights>Copyright A© Gannon J. Dick 2006-2008</dcterms:rights>
  </rdf:Description>
</rdf:RDF>
```