

RUST Privacy Advanced Methods

Introduction

The simple RUST behavior insures that PII and other sensitive data does not propagate to copies of the document in the formats chosen. In an alternate formulation, simple RUST behavior can be explained as an encipher by default, and exhibits well known, desirable cryptographic design properties.

1. The system must be practically, if not mathematically, indecipherable;
2. It must not be required to be secret, and it must be able to fall into the hands of the enemy without inconvenience;
3. Its key must be communicable and retainable without the help of written notes, and changeable or modifiable at the will of the correspondents;
4. It must be applicable to telegraphic correspondence;
5. It must be portable, and its usage and function must not require the concourse of several people;
6. Finally, it is necessary, given the circumstances that command its application, that the system be easy to use, requiring neither mental strain nor the knowledge of a long series of rules to observe.

-- [Kerckhoffs' principle](#) [military cryptography]

The Advanced Methods seek to compromise the default behavior of the system, true enough, but they are computational chains and no direct path exists to expose private information in a copy except as specified by the source. The “key” is distributed across several global properties of the document, so that decipherment will always require an exact copy of the source document and perhaps small modifications to that document. Decipherment, in short, always requires an overt act separate from sanitization of the information contained in the document.

Methods

1. Choice of Format

XHTML 1.0 with all meta data encoded with namespace qualified Dublin Core conventions. The only “required” meta data in an HTML/XHTML is a document title. The title (only) is duplicated according to syntax. RUST Privacy depends upon the existence of one single source file with the RUST behavior automatic in copies. The Advanced Methods require a target file (as meta data properties, ODF.base and ODF.filename).

2. Choice of Reference

The reference link (or anchor) is specified in the “publisher” field of a Bibliography Entry. In the case where any one of (*title,base,filename*) are missing, default RUST behavior is observed.

- i. Publisher = 'HTML' A link to a human-readable definition of the identifier or tag
- ii. Publisher = 'RDF' A link to a machine processable definition of the identifier or tag
- iii. Publisher = 'PII' An anchor, which includes original content.
- iv. Publisher = *null*, other Plain Text, a normal bibliographic mark (Short Name). Default RUST behavior.

3. Further Processing – Catalogs, Collations

XHTML documents with RDF type links can be rewritten in an XML format called a Resource Description Framework using a semi-standardized protocol called a [GRDDL Transform](#). This extract provides access to meta data definitions in a standardized way to enable development of an automated catalog system. Note that neither the text source of the XHTML nor the text source of the resultant GRDDL Transform contain original content of PII, only the (redacted form) class name proxy. An explanation of the output of the GRDDL Transform vis-a-vis the DCMI Abstract Model (DCAM) is in a [supplement to this document](#).

4. Further Processing – Redact in the usual way.

XHTML documents are XML and can be processed with general transforms. The “usual” way of redaction is to substitute a character, for example an “X”, for each character of the information being redacted. The PII text may also be replaced by an icon. In the case where Publisher = 'PII', only the PII content is transformed.

5. Further Processing – Redaction By Class Name Analysis

To a limited extent, documents with redactions which retain the class name are subject to analysis and a form of decryption, [Anchor•Buoy•Boat](#) Diagrams.

Results

The “advanced” directory of examples shows the results of the OpenDocumentText format to XHTML transformation (Methods 1 & 2) [Mayflower.xhtml], as well as the result of further processing. [Mayflower.rdf.xml] (Method 3) and [Mayflower-ruw.html] (Method 4).

Examples

Note: Examples use the original schema. A new schema is under development based upon the DCMI Abstract Model.

Supplement: [PII Resource Descriptions and the DCMI Abstract Model](#)

Discussion

The object of the Advanced Methods is not to expose PII or sensitive information but rather to design export tools which must be applied in serial fashion. While we are used to fundamental issues stated in terms of binary logic, the “transformation” of information is necessarily something more complex. The four results of the publisher field switch, for example, could be described as faithful/unfaithful for the reproduction of original content and same/different for the reference they make.

- i. Publisher = 'HTML' (unfaithful/different)
- ii. Publisher = 'RDF' (unfaithful/different)
- iii. Publisher = 'PII' (faithful/same)
- iv. Publisher = *null*, other (faithful/different)

The chain rule is simple, or maybe it just coalesces that wayⁱ; once unfaithful always unfaithful; once different always different. In this way the catalog record depends upon the agreement of both the author of the source and the tool maker programmer. In any case the author has the option of thwarting the tool maker, short of theft and manipulation of the source code. The monetization of traditional encryption tools injects a retailing principle at odds with good security. With RUST, there is no better version of encryption, costing more of course, to be available next year.

The "natural state" of Personally Identifiable Information is hidden, as is said, in plain sight of the author, who would seek to protect it by some encryption scheme. So, Kerckhoffs' Principles should apply to any copy of the document, regardless of who makes the copy.

Kerckhoffs**RUST**

The system must be practically, if not mathematically, indecipherable;

Once redacted, information is practically irretrievable. Forensic examination techniques of storage media with "memory" have come a long way since 1883. Still, once redacted *and moved to fresh media*, the information becomes mathematically indecipherable.

It must not be required to be secret, and it must be able to fall into the hands of the enemy without inconvenience;

Copies of the document do not enlighten any recipient as to the original private content. Sub-categories are public. [Public Categories of PII](#)

Its key must be communicable and retainable without the help of written notes, and changeable or modifiable at the will of the correspondents;

The "key" in this case is distributed switches affecting the content of the document copy.

It must be applicable to telegraphic correspondence;

Based upon the problems William Thompson (Lord Kelvin) encountered building the trans-Atlantic cable at about the same time, we should read this somewhat differently. As in the first principle media must not have "memory", a code must be amenable to simultaneous transmission, although we now know there is no such thing, the fastest transmission pipe extant.

It must be portable, and its usage and function must not require the concurrence of several people;

Not a fan of military "out-sourcing", we gather.

Finally, it is necessary, given the circumstances that command its application, that the system be easy to use, requiring neither mental strain nor the knowledge of a long series of rules to observe.

As FedEx put it, "So easy even the Chairman-Of-The-Board can do it.". The key to successful security is the motivation and the means to mind your own business.

- i Einstein did not disprove the common wisdom that "a stopped clock is right twice a day", rather he proved that a stopped clock on a fast rocket is right twice a day *at different times* than a stopped clock on earth. And, never ask an attractive Biologist for their phone number without also asking *when* it was their phone number. Rumor has it some phone numbers may have changed since the Jurassic period.