

RUST, The “Sequel”, Programmatic Redaction in Table Views

- RUST Transforms by SQL

- Copyright © 2008 RUSTPrivacy.org All Rights Reserved

Table of Contents

| | |
|--|----|
| RUST, The “Sequel”, Programmatic Redaction in Table Views..... | 1 |
| - RUST Transforms by SQL..... | 1 |
| Executive Summary..... | 2 |
| Introduction..... | 3 |
| Methods..... | 4 |
| Results..... | 6 |
| Conclusions..... | 9 |
| Appendix A..... | 11 |
| Appendix B..... | 12 |

Executive Summary

The RUST "Redact Unless Static Text" Rule, can be used to mask Personally Identifiable Information (PII) stored in a data base, and extracted to a data grid display. The redaction is column-wise. By restricting user modification privileges to SQL, data grid or view style reports can be easily tailored to audience without changes to selection criterion. This means that PII (without experimental - diagnostic - value) and Clinical Data (without e.g. billing value) can share the same grid without improper data disclosure. Nonetheless, in every case it is assumed that the entire grid will be visible to all users.

Introduction

Viewed from a meta data perspective, a data base is a sparsely populated Cartesian Grid. The spreadsheets used to record Financial and Experimental data give the impression of being densely populated because "uninteresting" meta data columns are never displayed. The problem of course is that "interesting" is a property of the audience, not of the data.

The sort of data breach which results from giving the wrong audience, even one considered friendly or trustworthy, something improperly interesting to read is easily avoidable.

Methods

Given this data grid:

| ID | Fname | Nname | Lname | SSN | Age | Gender |
|----|-----------|---------------|------------|-------------|-----|--------|
| 1 | George | Mr. President | Washington | 123-45-6789 | 67 | M |
| 2 | Martha | | Washington | 234-56-7890 | 42 | F |
| 3 | John | | Jones | 345-67-8901 | 53 | M |
| 4 | Mary | Spike | Jones | | | F |
| 5 | Alex | Bubba | Cohen | 156-45-9602 | 22 | F |
| 6 | William | Slick Willie | Slick | 235-85-0364 | | |
| 7 | Billy Bob | Tex | Bubba | | 15 | |

We can extract and sort the information, labeling it properly and getting rid of the Global Unique Identifier (GUID) which has no meaning outside of the domain (the computer on which the data is stored).

```
SELECT DISTINCT
"EG"."Fname" AS "Given Name",
"EG"."Nname" AS "Nick Name",
"EG"."Lname" AS "Family Name",
"EG"."SSN" AS "Social Security Number",
"EG"."Age" AS "Age",
"EG"."Gender" AS "Gender"
FROM "PII", "EG"
WHERE "PII"."_title" = 'null'
ORDER BY "EG"."Lname", "EG"."Fname"
```

The SQL is unremarkable, except perhaps for the inclusion of the PII table. The "null" row is exactly that, all nulls, and could be just as easily be entitled "dummy", "de nada" or "machts nichts". The DISTINCT grouping insures that the example (EG) values are not joined to all rows of the PII Table.

| Given Name | Nick Name | Family Name | Social Security Number | Age | Gender |
|------------|---------------|-------------|------------------------|-----|--------|
| Billy Bob | Tex | Bubba | | 15 | |
| Alex | Bubba | Cohen | 156-45-9602 | 22 | F |
| John | | Jones | 345-67-8901 | 53 | M |
| Mary | Spike | Jones | | | F |
| William | Slick Willie | Slick | 235-85-0364 | | |
| George | Mr. President | Washington | 123-45-6789 | 67 | M |
| Martha | | Washington | 234-56-7890 | 42 | F |

This is considered a wonderful domain independent result, but not, unfortunately, in the real world.

Results

Filling in the blanks

Filling in the blanks with generic terms, making the grid appear densely packed, may be desirable to one audience but undesirable for another. The Human Resources Department may like friendly Nick Names, but Mary Jones does not want to explain "Spike" any more than The Board wants the CFO called "Slick Willie".

```
SELECT DISTINCT
  "EG"."Fname" AS "Given Name",
  COALESCE( "EG"."Nname", 'Not Available' ) AS "Nick Name",
  "EG"."Lname" AS "Family Name",
  "EG"."SSN" AS "Social Security Number",
  "EG"."Age" AS "Age",
  "EG"."Gender" AS "Gender"
FROM "PII", "EG"
WHERE "PII"."_title" = 'null'
ORDER BY "EG"."Lname", "EG"."Fname"
```

| Given Name | Nick Name | Family Name | Social Security Number | Age | Gender |
|------------|---------------|-------------|------------------------|-----|--------|
| Billy Bob | Tex | Bubba | | 15 | |
| Alex | Bubba | Cohen | 156-45-9602 | 22 | F |
| John | Not Available | Jones | 345-67-8901 | 53 | M |
| Mary | Spike | Jones | | | F |
| William | Slick Willie | Slick | 235-85-0364 | | |
| George | Mr. President | Washington | 123-45-6789 | 67 | M |
| Martha | Not Available | Washington | 234-56-7890 | 42 | F |

One hopes that a bill collector looking for "Spike" Jones will find out how her husband "Not Available" Jones got his nick name. Oh, that's misleading isn't it ?

The COALESCE function is close, but not exactly what is needed.

Redaction

The proper way to dispense with the Nick Names is to redact them, make them blank *or something like it*. In doing so you should be careful that you do not inadvertently assign a "Blank" to unknown information, as this would mislead and increase disclosure liability^x. The PII name space provides the *something like it* in the form of a resource link. The hyper links work and point to a Resource Definition Framework (RDF) file. The links give a generic definition of what is meant by "Nick Name" or other generic baskets (see also Appendix A).

^x Personally Identifiable Information (PII) disclosure can depend on row selection outside of a full data grid (see also: Security and PII <http://www.RUSTPrivacy.org/PII.pdf>).

```

SELECT DISTINCT
"EG"."Fname" AS "Given Name",
CASE
    WHEN "EG"."Nname" is not null THEN "PII"."aka"
    ELSE 'Not Available'
END AS "Nick Name",
"EG"."Lname" AS "Family Name",
"EG"."SSN" AS "Social Security Number",
"EG"."Age" AS "Age",
"EG"."Gender" AS "Gender"
FROM "PII", "EG"
WHERE "PII"."_title" = 'about'
ORDER BY "EG"."Lname", "EG"."Fname"

```

| Given Name | Nick Name | Family Name | Social Security Number | Age | Gender |
|------------|---|-------------|------------------------|-----|--------|
| Billy Bob | http://purl.org/pii/terms/aka | Bubba | | 15 | |
| Alex | http://purl.org/pii/terms/aka | Cohen | 156-45-9602 | 22 | F |
| John | Not Available | Jones | 345-67-8901 | 53 | M |
| Mary | http://purl.org/pii/terms/aka | Jones | | | F |
| William | http://purl.org/pii/terms/aka | Slick | 235-85-0364 | | |
| George | http://purl.org/pii/terms/aka | Washington | 123-45-6789 | 67 | M |
| Martha | Not Available | Washington | 234-56-7890 | 42 | F |

Now the grid displays Nick Name properly, with the redaction accomplished only when the information was available in the data base. There are two other issues.

1. Assigning the correct PII "basket" to a column. For technical convention reasons we don't use the word "class"; that would mean Title Case terms and imply that the terms are differentiable further – a funny Nick Name, an embarrassing Nick Name or an obscene Nick Name, all of which exist but are not useful for the task at hand. A narrative explanation of the sixteen terms is at <http://purl.org/pii/terms/>. Social Security Number would be an "alpha" term, Age a "birth" term and Gender a "marks" term.
2. Data Base Columns which have a data type other than text should be converted to a text type. Most of the time the conversion is automatic, dates and times are the exception.

```

SELECT DISTINCT
"EG"."Fname" AS "Given Name",
CASE
    WHEN "EG"."Nname" is not null THEN "PII"."aka"
    ELSE 'Not Available'
END AS "Nick Name",
"EG"."Lname" AS "Family Name",

```

```

CASE
  WHEN "EG"."SSN" is not null THEN "PII"."alpha"
  ELSE 'Not Available'
END AS "Social Security Number",
CASE
  WHEN "EG"."Age" is not null THEN "PII"."birth"
  ELSE 'Not Available'
END AS "Age",
CASE
  WHEN "EG"."Gender" is not null THEN "PII"."marks"
  ELSE 'Not Available'
END AS "Gender"
FROM "PII", "EG"
WHERE "PII"."_title" = 'about'
ORDER BY "EG"."Lname", "EG"."Fname"

```

| Given Name | Nick Name | Family Name | Social Security Number | Age | Gender |
|------------|---|-------------|---|---|---|
| Billy Bob | http://purl.org/pii/terms/aka | Bubba | Not Available | http://purl.org/pii/terms/birth | Not Available |
| Alex | http://purl.org/pii/terms/aka | Cohen | http://purl.org/pii/terms/alpha | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/marks |
| John | Not Available | Jones | http://purl.org/pii/terms/alpha | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/marks |
| Mary | http://purl.org/pii/terms/aka | Jones | Not Available | Not Available | http://purl.org/pii/terms/marks |
| William | http://purl.org/pii/terms/aka | Slick | http://purl.org/pii/terms/alpha | Not Available | Not Available |
| George | http://purl.org/pii/terms/aka | Washington | http://purl.org/pii/terms/alpha | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/marks |
| Martha | Not Available | Washington | http://purl.org/pii/terms/alpha | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/marks |

It might be helpful to display "Not Available" as "PII"."misc" so that the columns are of a consistent type (a URI, link).

Conclusions

At this point, the grid can be tailored to audience:

Human Resources

| Given Name | Nick Name | Family Name | Social Security Number | Age | Gender |
|------------|---------------|-------------|------------------------|---|---|
| Billy Bob | Tex | Bubba | | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/misc |
| Alex | Bubba | Cohen | 156-45-9602 | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/marks |
| John | | Jones | 345-67-8901 | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/marks |
| Mary | Spike | Jones | | http://purl.org/pii/terms/misc | http://purl.org/pii/terms/marks |
| William | Slick Willie | Slick | 235-85-0364 | http://purl.org/pii/terms/misc | http://purl.org/pii/terms/misc |
| George | Mr. President | Washington | 123-45-6789 | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/marks |
| Martha | | Washington | 234-56-7890 | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/marks |

Finance

| Given Name | Nick Name | Family Name | Social Security Number | Age | Gender |
|------------|---|-------------|---|---|---|
| B. | http://purl.org/pii/terms/aka | Bubba | http://purl.org/pii/terms/misc | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/misc |
| A. | http://purl.org/pii/terms/aka | Cohen | http://purl.org/pii/terms/alpha | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/marks |
| J. | http://purl.org/pii/terms/misc | Jones | http://purl.org/pii/terms/alpha | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/marks |
| M. | http://purl.org/pii/terms/aka | Jones | http://purl.org/pii/terms/misc | http://purl.org/pii/terms/misc | http://purl.org/pii/terms/marks |
| W. | http://purl.org/pii/terms/aka | Slick | http://purl.org/pii/terms/alpha | http://purl.org/pii/terms/misc | http://purl.org/pii/terms/misc |
| G. | http://purl.org/pii/terms/aka | Washington | http://purl.org/pii/terms/alpha | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/marks |
| M. | http://purl.org/pii/terms/misc | Washington | http://purl.org/pii/terms/alpha | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/marks |

Clinic

| Given Name | Nick Name | Family Name | Social Security Number | Age | Gender |
|------------|---------------|-------------|---|-----|---------------|
| Billy Bob | Tex | B. | http://purl.org/pii/terms/misc | 15 | Not Available |
| Alex | Bubba | C. | http://purl.org/pii/terms/alpha | 22 | F |
| John | | J. | http://purl.org/pii/terms/alpha | 53 | M |
| Mary | Spike | J. | http://purl.org/pii/terms/misc | | F |
| William | Slick Willie | S. | http://purl.org/pii/terms/alpha | | Not Available |
| George | Mr. President | W. | http://purl.org/pii/terms/alpha | 67 | M |
| Martha | | W. | http://purl.org/pii/terms/alpha | 42 | F |

Waiting Room

| Given Name | Nick Name | Family Name | Social Security Number | Age | Gender |
|------------|-----------|-------------|---|---|---|
| B. | Tex | Bubba | http://purl.org/pii/terms/misc | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/misc |
| A. | Ms. | Cohen | http://purl.org/pii/terms/alpha | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/marks |
| J. | Mr. | Jones | http://purl.org/pii/terms/alpha | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/marks |
| M. | Ms. | Jones | http://purl.org/pii/terms/misc | http://purl.org/pii/terms/misc | http://purl.org/pii/terms/marks |
| W. | | Slick | http://purl.org/pii/terms/alpha | http://purl.org/pii/terms/misc | http://purl.org/pii/terms/misc |
| G. | Mr. | Washington | http://purl.org/pii/terms/alpha | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/marks |
| M. | Ms. | Washington | http://purl.org/pii/terms/alpha | http://purl.org/pii/terms/birth | http://purl.org/pii/terms/marks |

The SQL queries for each of these audiences are provided in Appendix B. In some cases they involve adjunct tricks such as term of address replacement in the Waiting Room example above – note that Age and Gender are *used*, but not disclosed.

Various levels of disclosure can be built into report grids programmatically. For further reading on RUST see <http://www.RUSTPrivacy.org/> .

Appendix A

A spreadsheet of the PII Table as well as the result grids is available at <http://www.RUSTPrivacy.org/2008/PIIeg.xls> .

The Persistent Uniform Resource Locator's (PURLs) for the PII name space lack a bit of user friendliness, but the XML Meta Data Profile (XMDP) links may be used too.

| (text) | PURL | Link |
|---------------|---|---|
| [aka] | http://purl.org/pii/aka | http://www.RUSTPrivacy.org/2008/pii-history.html#aka-003 |
| [alpha] | http://purl.org/pii/alpha | http://www.RUSTPrivacy.org/2008/pii-history.html#alpha-003 |
| [birth] | http://purl.org/pii/birth | http://www.RUSTPrivacy.org/2008/pii-history.html#birth-003 |
| [bystander] | http://purl.org/pii/bystander | http://www.RUSTPrivacy.org/2008/pii-history.html#bystander-003 |
| [car] | http://purl.org/pii/car | http://www.RUSTPrivacy.org/2008/pii-history.html#car-003 |
| [contact] | http://purl.org/pii/contact | http://www.RUSTPrivacy.org/2008/pii-history.html#contact-003 |
| [death] | http://purl.org/pii/death | http://www.RUSTPrivacy.org/2008/pii-history.html#death-003 |
| [dna] | http://purl.org/pii/dna | http://www.RUSTPrivacy.org/2008/pii-history.html#dna-003 |
| [groups] | http://purl.org/pii/groups | http://www.RUSTPrivacy.org/2008/pii-history.html#groups-003 |
| [health] | http://purl.org/pii/health | http://www.RUSTPrivacy.org/2008/pii-history.html#health-003 |
| [ice] | http://purl.org/pii/ice | http://www.RUSTPrivacy.org/2008/pii-history.html#ice-003 |
| [location] | http://purl.org/pii/location | http://www.RUSTPrivacy.org/2008/pii-history.html#location-003 |
| [marks] | http://purl.org/pii/marks | http://www.RUSTPrivacy.org/2008/pii-history.html#marks-003 |
| [misc] | http://purl.org/pii/misc | http://www.RUSTPrivacy.org/2008/pii-history.html#misc-002 |
| [money] | http://purl.org/pii/money | http://www.RUSTPrivacy.org/2008/pii-history.html#money-003 |
| [witness] | http://purl.org/pii/witness | http://www.RUSTPrivacy.org/2008/pii-history.html#witness-003 |

Both Human-Readable and Resource Encoded Links are available.

Appendix B

Human Resources

```
SELECT DISTINCT

"EG"."Fname" AS "Given Name",

"EG"."Nname" AS "Nick Name",

"EG"."Lname" AS "Family Name",

"EG"."SSN" AS "Social Security Number",

CASE WHEN "EG"."Age" is not null THEN "PII"."birth" ELSE "PII"."misc" END AS "Age",

CASE WHEN "EG"."Gender" is not null THEN "PII"."marks" ELSE "PII"."misc" END AS "Gender"

FROM "PII", "EG"

WHERE "PII"."_title" = 'about'

ORDER BY "EG"."Lname", "EG"."Fname"
```

Finance

```
SELECT DISTINCT

CONCAT(LEFT("EG"."Fname", 1), '.') AS "Given Name",

CASE WHEN "EG"."Nname" is not null THEN "PII"."aka" ELSE "PII"."misc" END AS "Nick Name",

"EG"."Lname" AS "Family Name",

CASE WHEN "EG"."SSN" is not null THEN "PII"."alpha" ELSE "PII"."misc" END AS "Social Security Number",

CASE WHEN "EG"."Age" is not null THEN "PII"."birth" ELSE "PII"."misc" END AS "Age",

CASE WHEN "EG"."Gender" is not null THEN "PII"."marks" ELSE "PII"."misc" END AS "Gender"

FROM "PII", "EG"

WHERE "PII"."_title" = 'about'

ORDER BY "EG"."Lname", LEFT("EG"."Fname", 1)
```

Clinic

```
SELECT DISTINCT

"EG"."Fname" AS "Given Name",

"EG"."Nname" AS "Nick Name",

CONCAT(LEFT("EG"."Lname",1), '.') AS "Family Name",

CASE WHEN "EG"."SSN" is not null THEN "PII"."alpha" ELSE "PII"."misc" END AS "Social Security Number",

"EG"."Age" AS "Age",

CASE WHEN "EG"."Gender" is not null THEN "EG"."Gender" ELSE 'Not Available' END AS "Gender"

FROM "PII", "EG"

WHERE "PII"."_title" = 'about'

ORDER BY LEFT("EG"."Lname", 1), "EG"."Fname"
```

Waiting Room

```
SELECT DISTINCT

CONCAT(LEFT("EG"."Fname",1), '.') AS "Given Name",

CASE

    WHEN "EG"."Age" <21 THEN "EG"."Nname"

    WHEN "EG"."Gender" = 'M' THEN 'Mr.'

    WHEN "EG"."Gender" = 'F' THEN 'Ms.'

    ELSE ""

END AS "Nick Name",

"EG"."Lname" AS "Family Name",

CASE WHEN "EG"."SSN" is not null THEN "PII"."alpha" ELSE "PII"."misc" END AS "Social Security Number",

CASE WHEN "EG"."Age" is not null THEN "PII"."birth" ELSE "PII"."misc" END AS "Age",

CASE WHEN "EG"."Gender" is not null THEN "PII"."marks" ELSE "PII"."misc" END AS "Gender"

FROM "PII", "EG"

WHERE "PII"."_title" = 'about'

ORDER BY "EG"."Lname", LEFT("EG"."Fname",1)
```