A DEIDENTIFICATION APPROACH TO DNT

A Path Forward to Creating a W3C DNT Standard

What is Do Not Track?

- **Tracking:** Tracking is the collection, use, or retention of data records across non-affiliated websites that are, or can be, associated with a specific user, user agent, or device.
  - In essence – the historical record of a user's visits across websites

- To "Not Track" we need to remove the historical record of a user's online activity
  - Unique ID + Activity (URL) = Historical Record

- In a simplified model, this can be achieved in two ways:
  - Remove all unique IDs tied to a specific device, or
  - Remove all of the URLs tied to a specific device

How to Not Track

- **Removing Unique IDs**
  - One method would be to aggregate away all unique IDs (346 visitors viewed the http://ourhomepage.com on Jan 1, 2010)
  - Another method is to transform IDs such that they no longer associate to an operationally specific device (cookie ID B8Y is available for reporting but no cookie ID with this value exists in the real world)

- **Removing URLs**
  - One method would be to aggregate a value representing many different URLs such as “aggregate scoring” (cookie ID 123 has a score of 12 for “online vehicle interest”)

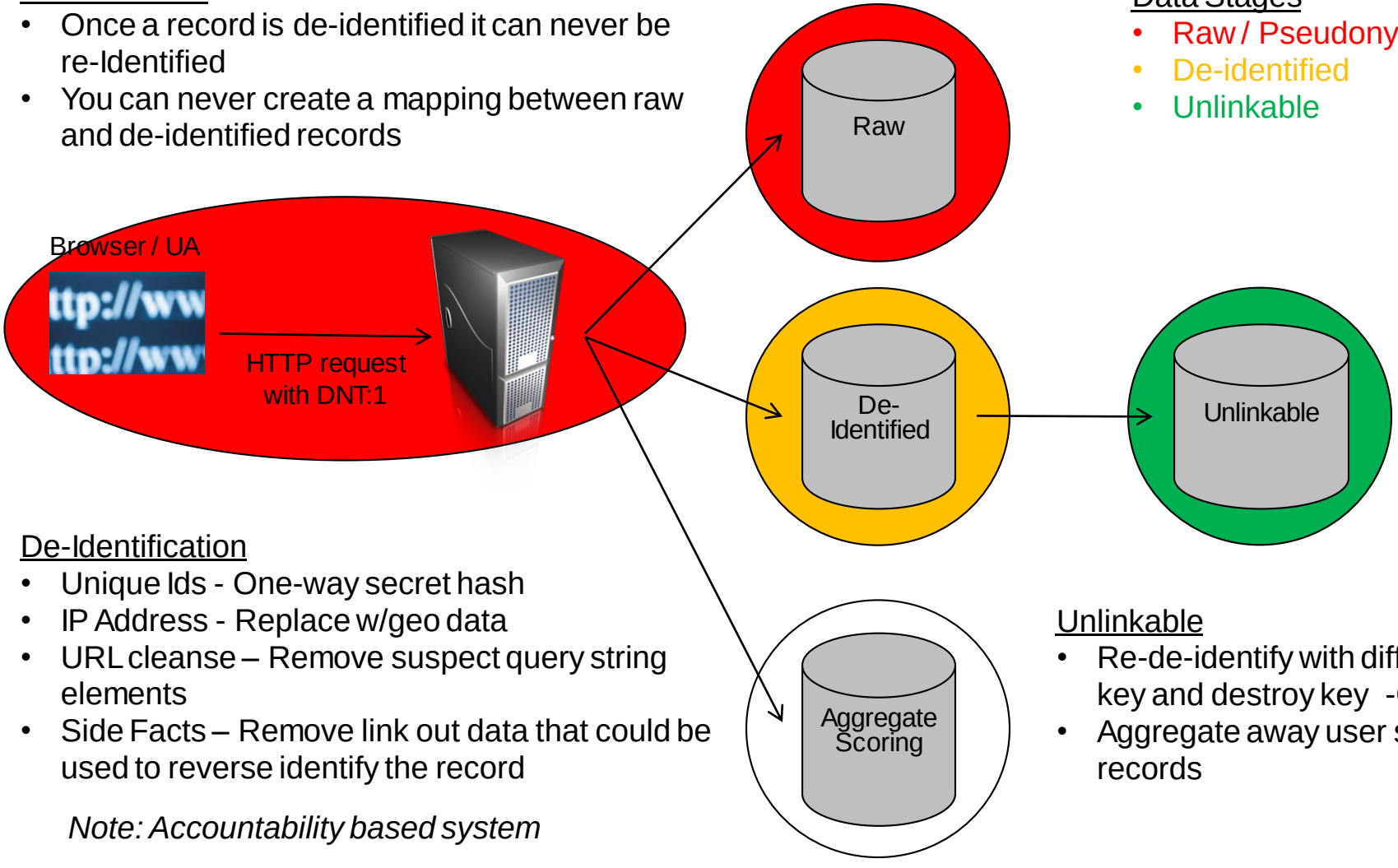
- In both cases, the tie between an operational unique ID and a URL (activity) has been removed. A specific device’s historical trail no longer exists.

Golden Rules

- Once a record is de-identified it can never be re-Identified
- You can never create a mapping between raw and de-identified records

Data Stages

- Raw / Pseudonymous
- De-identified
- Unlinkable



De-Identification

- Unique Ids - One-way secret hash
- IP Address - Replace w/geo data
- URL cleanse – Remove suspect query string elements
- Side Facts – Remove link out data that could be used to reverse identify the record

Note: Accountability based system

Unlinkable

- Re-de-identify with different key and destroy key -OR-
- Aggregate away user specific records

Permitted Uses in the Tri-State

- Raw Data / Pseudonymous Data
  - **Permitted Uses:** Security/Fraud, Frequency Capping, Debugging, (some) Financial/Audit
  - **Principle:** Areas where an operational ID is needed.
  - **Goal:** Define shorter retention timeframes where possible.

- De-Identified Data
  - **Permitted Uses:** Financial/Audit, Product Improvement, Market Research
  - **Approach:** One-way secret hash, operational/administrative controls.
  - **Goal:** Data is not able to be used for production purposes - unable to alter a specific user's online experience.

- Unlinked Data
  - **Permitted Uses:** Any use (out of scope of DNT at this point)
  - **Approaches:** Re-one-way secret, further data minimization, and destroy key – and/or– aggregation to remove user specific records
  - **Goal:** Data should be able to be freely shared with 'no' risk of re-identification

Providing More Choices

- As there are two approaches to removing a user's historical activity from data records, it may be counter-intuitive that a profile could still be created of a user's interests without retaining any record of their activity ("aggregate scoring")

- **AdChoices**
  - By combining Do Not Track with AdChoices we bring together a more powerful combination to empower users with greater choices and flexibility
  - Do Not Track = no historical activity tied to a specific operational device
  - AdChoices = no profiling or behavioral targeting to a specific operational device

- By leveraging the benefits of both programs, we give users the option to receive personalized content and ads and not have their historical records retained at the same time. Of course, users can turn off both if they like.

+

Questions?