

DATA MANAGEMENT WHITE PAPER SERIES: Paper 1. Draft – November 1st, 2013

Master Data Management and HUD's Need for Unique IDs

Goal: Minimize use and system replication of SSN and other Personally Identifying Information

Summary: This White Paper proposes that HUD's Master Data Management (MDM) system create a unique ID for each HUD client and partner and that unique ID replace the use of SSN, EIN, and TIN as the primary key in HUD's systems. The PII related to the HUD unique ID would be maintained in a single MDM database accessible as a look-up for external partners to obtain the unique ID and for other business reasons to support program operations. HUD's other systems would no longer maintain the PII data. This would serve the dual goal of reducing replication of personal data and improving HUD's ability to report on program overlap.

Background. Public Housing Authorities, owners of project based rental assistance properties, local governments, lenders, businesses, and other service providers provide HUD with significant amounts of personal data about themselves and the individuals participating in our programs. The data is required for our program operations but it is possible to substantially reduce the number of systems containing the information.

Many of HUD's internal databases use the client's social security number (SSN) as the **primary key** to access the data and link database tables. Similarly, business taxpayer identity numbers (TIN) or employer identity numbers (EIN) are used as primary keys for accessing data about HUD's partners. Due to the increasing need to protect the personally identifiable information (PII) about HUD's clients and partners (tenants, homebuyers, owners, managers, etc.) widespread use of SSNs, TINs and EINs is no longer acceptable. HUD cannot fully remove its reliance on these numbers primarily because HUD relies on data from 3rd parties and these numerical identifiers help ensure HUD receives the appropriate information for the appropriate person or entity.

Proposal.

The first issue that needs to be addressed with SSN is that there are a finite number of combinations and any mathematical or replacement routine would be easily circumvented using a brute force approach to resolve. The second issue is the translation from one SSN to unique ID or vice versa needs to be tightly controlled; otherwise it would be possible to call the conversion routine to create a lookup table for every possible SSN. The third issue is that users should not confuse SSN with the unique ID, therefore the unique ID needs to be non-numeric.

The advantage of only a finite number of SSN/TIN/EIN, it is possible to pre-generate ever possible unique ID prior to the conversion of applications to using unique ID. This would cut down on the complexity of handling conversions, as unique ID would not have to be generated for any new SSN.

Unique Id would be two parts, a random combination of eight alpha characters, removing those characters that could be confused with numbers e.g. I for one or O for zero. The last character would be a check digit (http://en.wikipedia.org/wiki/Check_digit). This would allow applications to easily validate an unique ID prior to submission for conversion.

e.g 123-45-6789 translates into FGE-JW-EEGF where F is the check digit

TIN, EIN and SSN with the value would not translate to the same unique ID. Each unique ID would only be used for one conversion, not SSN and EIN or TIN and EIN.

Access to the MDM database would have to be strictly controlled and monitored. Applications would therefore convert from SSN/TIN/EIN to the unique ID and back using service-oriented architecture (SOA). These services can be provided on a variety of platforms to support the differing needs of the Departments application and the platforms those applications are using. For web based applications it would be possible to provide conversion using Web Services. For database access it would be possible to provide stored procedure access. For text files, services would provide generic bulk translations. Migrating to unique ID would also not have to be a 'big bang' approach. Applications could migrate as services become available and as funding for the conversion becomes available.

The ultimate goal for this effort is that no application stores SSN/TIN/EIN is part of their database. SSN information transferred from one application to another uses unique keys. Only when an application is providing the information to our business partners, SSA, Treasury, Public Housing Authority, is the information converted to a real SSN.

	Alternate Key		
	SSN	TIN	EIN
000-00-0000	BMP-BB-DKUZ	CRD-FX-ZAWD	RSM-HH-XYBF
000-00-0001	LNU-HL-WEBM	XZA-PM-MPFF	FGA-RD-MPLB
...			
457-12-9869	FEJ-WM-QMKE	RDL-FW-RMBU	TYT-MB-UFSW
...			
999-99-9999	ARB-BW-FQDB	PMH-FE-WMPG	LTM-GG-RDWM

Sample Application Process Flow

