7. Impact of VAWA on UID technologies

In January 2006, Congress passed The Violence Against Women and Department of Justice Reauthorization Act of 2005, H.R. 3402 ("VAWA")[10], which has a profound impact on HMIS data elements and on protecting against the previously described privacy threats. VAWA makes special provisions for a HMIS to use UIDs. Specifically, section 605 (A) states that "Victim service providers … shall… not disclose for purposes of a … [HMIS] personally identifying information about any client. …The Secretary may … require … for purposes of HMIS non-personally identifying data that has been de-identified, encrypted, or otherwise encoded.…"

VAWA defines the phrase "personally identifying information" in 605 (A) as:

> PERSONALLY IDENTIFYING INFORMATION OR PERSONAL INFORMATION.— The term 'personally identifying information' or 'personal information' means individually identifying information for or about an individual including information likely to disclose the location of a victim of domestic violence, dating violence, sexual assault, or stalking, including— ''(I) a first and last name; ''(II) a home or other physical address; ''(III) contact information (including a postal, e-mail or Internet protocol address, or telephone or facsimile number); ''(IV) a social security number; and ''(V) any other information, including date of birth, racial or ethnic background, or religious affiliation, that, in combination with any other non-personally identifying information would serve to identify any individual.

VAWA effects HMIS in two significant ways. First, VAWA supports using a UID instead of explicit identifiers. Second, VAWA requires a HMIS to use a set of data elements and a technology for processing UIDs such that no Client can be re-identified. Prior to VAWA, HUD had been following the pattern of recent U.S. privacy regulation in which technologies combine with practices and policies to provide a minimal risk of re-identification.[11] The wording of VAWA, however, insists on guaranteed protection against re-identification.

[10] This was a reauthorization of the earlier VAWA (of 1998 and then 2000). It continues to focus on ending domestic violence, sexual assault, dating violence and stalking. It sets priorities and funding levels, determines options available to victims of abuse, sets criminal justice system responses to violence, and establishes national investments in prevention. Special considerations are given to HMIS under Title VI (Housing Opportunities and Safety for Battered Women and Children), Subtitle N (Addressing the Housing Needs of Victims of Domestic Violence, Dating Violence, Sexual Assault, and Stalking), Section 605, Amendment to Section 423 of the McKinney-Vento Homeless Assistance Act (42 U.S.C. 11383).

[11] In recent U.S. privacy legislation, the notion of minimal risk of re-identification appears in the medical privacy regulation known as HIPAA (Health Information Portability and Accountability Act). Under HIPAA, 45 C.F.R. § 164.514 (b)(1)(2002), patient health data may be shared outside the patient's care if "the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information." This analysis must be based on generally accepted statistical and scientific principles, and the person who makes this finding must have "appropriate knowledge and experience applying generally accepted statistical and scientific principles and methods for rendering information not individually identifiable." Unlike HIPAA, VAWA's wording is not of minimal risk but of guaranteed protection against re-identification.

7.1 VAWA's impact on data elements

VAWA requires changes in several fields in the Dataset because some of the fields enable the Dataset to link to other available information to re-identify Clients. HUD will announce a revision to the Dataset shortly, but the following recommendations reflect relevant observations made earlier in this writing.

Recommendation #18: The fields date of birth*,* gender*, and* ZIP code of last residence *must contain less specific information than the full month, day, and year of birth, and all 5 digits of the ZIP code.*

As was shown in Figure 13, these values currently allow re-identifications, but using more general values such as *age* and the *first 3 digits of the ZIP code* significantly reduces re-identifications without affecting the utility of the information in the AHAR.

Modifications will probably exclude the PIN from the Dataset in order to limit unnecessary risk of linking the Dataset to other non-HMIS data released from the same Shelter. The *ethnicity* and *race* fields may require special handling. The Program-Specific Data Elements require additional consideration in light of other kinds of data from social service programs that a Planning Office may hold. This vulnerability may differ among municipalities and states as different kinds of secondary data from related programs are available.

Section 8 introduces PrivaMix as a UID technology hat meets the higher privacy standard established by VAWA. Section 10 gives utility and privacy results when PrivaMix was used in a real-world experiment in Iowa. Section 11 then re-examines the identifiability of HMIS data elements in light of VAWA and PrivaMix.

7.2 VAWA's impact on initial UID technologies

UID technologies that Planning Offices had previously explored for constructing, maintaining and using UIDs now face additional hurdles with the passage of VAWA. See Section 6 for an assesssment of these technologies pre-VAWA. The following subsections describe the additional difficulties faced by these technologies in attempting to comply with the privacy standard established by VAWA.

7.2.1. Consent (not under VAWA)

"Consent" as a UID technology refers to a permission technology. The database that stores Client information at Shelters includes a permission flag that records whether a Client has granted permission to have her data forwarded to a Planning Office. The Planning Office only receives the information of Clients who granted permission. If the Planning Office receives explicitly identified UIDs, such as name and Social Security numbers, from consenting Clients, VAWA does not allow this approach. (Section 6.6 describes Consent as UID technology in more detail.)

7.2.2. Biometrics (not under VAWA)

Using a biometric as source information for a UID technology has the advantage that the biometric is something always present with the Client and that typically does not change. The most common biometric is a fingerprint. It can be source information to a hash or encryption function or the fingerprint itself can be the UID. VAWA prohibits fingerprint-based UIDs if fingerprint data can match law-enforcement data. (Section 6.5 describes the use of biometrics as UID technology in more detail.)

7.2.3. Scan Cards / RFID tags (maybe okay under VAWA)

Using Scan Cards as a UID technology involves issuing a card containing a UID to each Client who presents for service. Scan cards that have a magnetic strip on one side resemble credit cards. Information is stored on the magnetic strip. Radio frequency identification (RFID) cards have no magnetic strip. Information is stored within the card. A card reader can read the information even though it is not visible to the human eye. Magnetic strips are typically readable by most card readers, not just those at the issuing Shelter. A big downside to using scan cards is not from VAWA but from practical matters such as handling lost, swapped or stolen cards.

The decision of what information appears on the card determines its acceptability under VAWA. A randomly assigned number at each Shelter is fine, but VAWA may not allow the card to include explicit identifiers and certain demographics. If Shelters share the same card and the number associated with the card appears in other data, then privacy threats may exist. (Section 6.4 describes the use of scan cards as a UID technology in more detail.)

7.2.4. Encoding (problematical under VAWA)

Using "encoding" to produce UIDs simply involves concatenating parts of source information to form a UID. De-duplication is then performed by simply matching resulting UID values. An obvious problem with encoding is that given a series of UIDs and some source information, an attacker can often deduce what parts of which source information appears in the UID. When the source information uses demographics and explicit identifiers, the encoding may be problematical under VAWA. (Section 6.1 describes encoding as a UID technology in more detail.)

7.2.5. Hashing (problematical under VAWA)

Using "hashing" to produce UIDs involves computing a number from source information. A vendor can create a hash function, but if someone is highly motivated, he can often reverse an ad hoc approach. Protection using an ad hoc hash function is good only as long as no one learns the actual hash function used. Rather than using ad hoc hash functions, cryptographically strong or one-way hash methods are highly recommended. With a strong hash function, everyone can examine the method used, but still cannot reverse the process without performing more computation than is feasible. (Section 6.2 describes hashing as a UID technology in more detail.)

If the same hash values are broadly used with Clients, then they may lead to re-identifications through linking. If the intimate stalker compromises a Shelter or the Planning Office, he can learn a targeted Client's hashed UID and use it to locate the targeted Client. Further, if the source information is a SSN or demographics, then the Planning Office, with access to the hash function and knowledge of the kind of source information used, could re-identify all UIDs by exhaustively computing all UIDs (a dictionary attack). For these reasons, hashing, even strong hashing is problematical under VAWA.

7.2.6. Encryption (problematical under VAWA)

Using encryption as a UID method is similar to hashing except with encryption there exists a "key" such that whoever has the key can reverse the process to take a UID and reveal some (or all) of the source information that produced it. The discussion and shortcomings are the same as for hashing (above in Section 7.2.5), including VAWA concerns, with the additional consideration that only the Shelter may be able to hold the key. (Section 6.3 describes encryption as a UID technology in more detail.)

The following recommendation summarizes the findings.

<u>Recommendation #19:</u> The technology used to construct and de-duplicate UIDs must satisfy VAWA's requirements limiting re-identification. Consent and biometrics appear unable to satisfy the privacy standard established by VAWA. Encoding, hashing, and encryption may enable unwanted linking, and if so, pose grave concerns in attempts to use them to satisfy VAWA's privacy standard. Scan cards and RFID tags may be used, depending on the information appearing on (or within) the card.

Of all the UID technologies assessed in Section 6, inconsistent hashing and distributed query has the best privacy results, so it is not surprising that they form the basis for PrivaMix, which is described in the next section as a UID technology that satisfies the higher privacy standard imposed by VAWA.