RSR in Focus

Client De-Duplication: How to Avoid Upload Errors

The RSR Web System **does not accept** client-level data files with duplicate eUCIs.

The Ryan White Services Report (RSR) requires that providers submit client-level demographic, service, and clinical data for Ryan White HIV/AIDS Program clients. To maintain client anonymity, RSR client-level data do not include any personally identifiable information, such as name, Social Security Number or address. Records are identified only through an encrypted Unique Client Identifier (eUCI). This eUCI protects each client's identity and ensures the privacy of personal health information. It also allows the HIV/AIDS Bureau (HAB) to link data that belong to the same client across multiple providers to de-duplicate client counts.

What Makes up the eUCI?

The eUCI is constructed in four steps:

1. An **11-digit UCI** is developed from:
   - 1st and 3rd characters of first and last names
   - Full date of birth: MM/DD/YY
   - Gender code
2. Within a given data system, a **suffix to indicate UCI uniqueness** is added. If only one client within a site has a given UCI, the suffix is "U" for unique. If more than one client has the same UCI, the final character of the first client's UCI is "A," the final character of the second client's UCI is "B," and so on. The suffix limits the number of clients who have the same eUCI.
3. The 11-digit UCI is **encrypted** with the SHA-1 algorithm to create a 40-character eUCI.
4. The "suffix to indicate UCI uniqueness" is added as the **41st character**.

How Can I De-duplicate My Data?

Unfortunately, because the underlying data elements of the eUCI are not unique, multiple clients may share the same 11-digit UCI. Although the eUCI allows providers to distinguish between these clients by using the 41st character, some providers have uploaded client-level data XML files with multiple records with the same eUCI. The RSR Web System does not accept these files.

Providers can use several mechanisms for de-duplicating their data:

This resource was prepared by CAI Global and their partners Abt Associates and Mission Analytics under Cooperative Agreement # UF2HA26520 from the Health Resources and Services Administration's HIV/AIDS Bureau. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the HIV/AIDS Bureau.

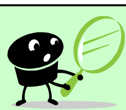
- **After submission:** If providers receive a de-duplication validation error at upload, they should identify which records in the client-level data XML file share the same eUCI. Then, they can go back to their source systems to review the records' other data, such as first and last name, address and race, to determine if these records actually belong to the same person. If so, one record should be appropriately updated, while the other record should be deleted. If they don't belong to the same person, the provider must assign one record an "A" and the other record a "B" for the 41st character. The process of adding this digit will depend on the system you use.

Overview of De-duplication Process

Providers are expected to determine whether records with the same eUCI actually belong to the same person.

- If so, records should be merged.
- If not, the provider should change the 41st digit to "A", "B", etc., to distinguish these clients.

- **Continual de-duplication by system administrator:** Providers or systems administrators that manage networked provider systems may run periodic checks throughout the reporting period on data to identify duplicate records. In this case, you may need to help by reviewing identifying data elements, as above. Once again, the provider or system administrator would need to merge records or assign the 41st character once potential duplicates are found.



How Do I Know If Records Belong to the Same Client?

Check other identifying data elements, such as:

- First name
- Last name
- Ethnicity
- Race
- Address
- Phone number

This process often involves human review to identify matches that are "close enough," which account for nicknames, changes in address, etc.

- **At the point of data entry:** Ideally, de-duplication would occur when the provider is creating a new and potentially duplicate record. When a provider creates a new record for a client, they would get a flag if the UCI already exists in the database. The provider would then decide whether to create a new record or add to the original one. Check to see if you can build this feature into your system and workflow.

This resource was prepared by CAI Global and their partners Abt Associates and Mission Analytics under Cooperative Agreement # UF2HA26520 from the Health Resources and Services Administration's HIV/AIDS Bureau. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the HIV/AIDS Bureau.