L - 21: De-Identifying Protected Information and Limited Data Sets

POLICY: See HHSA-L-21 De-Identifying Protected Information and Limited Data Sets, at www.cosdcompliance.org.

DEFINITIONS: See HHSA Policy L-30 Privacy Definitions.

PROCEDURES:

1. **De-Identified Data** – When an appropriate business need has been determined, HHSA programs may disclose data that does not identify an individual and for which there is no reasonable basis to believe that it could be used to identify an individual (i.e. de-identified data). HHSA programs must use one of the two methods below to ensure data is de-identified:

   a. **Safe Harbor Method** – a program may assume data is de-identified if the program does not have knowledge that the information could be used alone, or in combination with other information, to identify an individual *and* each of the following identifiers are removed:

      i. Names, including first, last and initials

      ii. All geographic subdivisions smaller than a state, including street address, city, precinct, zip code, and their equivalent geocodes; except for the initial three digits of a zip code if, according to the current Bureau of the Census:

         a. The geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people; and

         b. The initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to "000"

      iii. All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death

      iv. All ages over 89 and all elements of dates (*including year*) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older

      v. Telephone and fax numbers

      vi. Email addresses, URLs, and IP addresses

      vii. Social security numbers (even the last 4 digits), medical record numbers, and health plan beneficiary numbers

      viii. Any type of account numbers

      ix. Certificate or license numbers

      x. Vehicle identifiers and serial numbers, including license plate numbers

      xi. Biometric identifiers, including finger and voice prints; as well as device identifiers and device serial numbers

      xii. Full face photographic images and any comparable images

xiii. Any other unique identifying number, characteristic, or code; except as permitted in the Record Identifiers section of this document

b. **Expert Determination** – if a program wishes to include some fields that are excluded from the Safe Harbor Method described in this document, the program may request an expert determination from Business Assurance and Compliance (BAC).

   i. The program will provide list of fields, sample data, or the actual data set to BAC; as well as any additional information or context requested by BAC.

   ii. BAC will evaluate the extent to which the information can (or cannot) be identified.

      a. If the risk of identification is determined to be *very small*, then the data may be considered de-identified.

      b. If the risk of identification is determined to be *greater than very small*, then BAC will provide guidance as to what, if any, statistical or scientific methods can be applied to mitigate the anticipated risk, such as to remove, roll-up, redact, or mask data.

         - After the program has made changes as suggested by BAC, the program will resubmit the data and BAC will re-evaluate the resulting data to confirm the risk is no more than very small. If the risk is again determined to be greater than very small, BAC will provide additional guidance.

      c. BAC will document the methods and results of the analysis that justify the determination.

2. **Limited Data Sets** – HHSA programs may use or disclose certain specific data (i.e. limited data set) for the purposes of research or public health if *all* the following are met:

   a. HHSA enters into a Data Use Agreement (DUA), Business Associate Agreement (BAA), or similar agreement with the recipient. The DUA, BAA, or similar agreement must be approved by BAC.

   b. The limited data set disclosed *excludes all* the following identifiers:

      i. Names

      ii. Postal address information (i.e. street or PO Box address)

      iii. Telephone and fax numbers

      iv. Email addresses, URLs, and IP address

      v. Social security, medical record, or health plan beneficiary numbers

      vi. Account numbers

      vii. Certificate or license numbers

      viii. Vehicle identifiers and serial numbers, including license plate numbers

      ix. Biometric identifiers, including finger and voice prints

      x. Device identifiers and device serial numbers

      xi. Full face photographic images and any comparable images.

   c. The data disclosed may include the following identifiers:

      i. All elements of dates

ii. Geographic elements of town, city, and full zip code

iii. Record identifiers as described in the Record Identifiers section of this document

3. **Record Identifiers** – HHSA programs may assign a code or other means of record identification (i.e. record identifier) to allow HHSA programs a means of re-identifying data, provided that the following conditions are met.

   a. Derivation – The record identifier is not derived from or related to information about the individual and is not otherwise capable of being translated as to identify the individual.

   b. Security – HHSA does not use or disclose the record identifier for any other purpose and does not disclose the mechanism for re-identification.

QUESTIONS/INFORMATION: Please contact HHSA Business Assurance and Compliance (BAC) by email at Compliance.HHSA@SDCounty.ca.gov or by phone (619) 237-8571.