

April 16, 2018

RE: Data Confidentiality and Security Standards for APPD LEARN's de-identification system at learn.cybermango.org

Programs participating in APPD LEARN studies provide their research data identified by an APPD LEARN id, rather than personal identifiers. This letter explains how the APPD LEARN id is generated, describes APPD LEARN's online implementation of the de-identification system at learn.cybermango.org, details the provisions implemented in the learn.cybermango.org system to protect data confidentiality, and affirms APPD LEARN's commitment to maintaining such confidentiality.

1. The APPD LEARN id

APPD LEARN ids are generated by applying a one-way encryption algorithm to participant date of birth and last four digits of social security number. The resulting id is consistent, but can not be decrypted to reidentify the birthdate or ssn4 components. Programs in APPD LEARN studies transmit their data using these APPD LEARN ids, rather than any personal identifiers—APPD LEARN does not receive the birthdate or ssn4.

As a further confidentiality measure, prior to sharing any data set in APPD LEARN, the APPD LEARN ids are encrypted a second time with a one-way algorithm that also incorporates a secret key known only to APPD LEARN. This prevents even the programs submitting data from re-identifying participants based on the id.

2. Connection to learn.cybermango.org

To simplify the generation of correct APPD LEARN ids, APPD LEARN provides a web-based de-identification system at learn.cybermango.org, a password-protected Linux server housed in a locked office at the University of Illinois at Chicago. All hard drive partitions on the server, including virtual memory (swap space) are encrypted.

learn.cybermango.org accepts SSL-encrypted connections from browser to the de-identification system.

3. Metadata elements

The APPD LEARN de-identification system does not retain or log any of the uploaded personal information from which it creates APPD LEARN ids.



Browser IP addresses for all accesses to any site on the web server are routinely logged as a security measure (e.g. to disable access to the site by a client attempting attacks on the server), but these logs do not contain any participant data.

The APPD LEARN deidentification system at learn.cybermang.org does not use third-party cookies or embedded analytics (such as Google Analytics) that would enable tracking participants to other sites or identification of participation to other sites.

4. Data storage

No uploaded data is stored as a result of interaction with the APPD LEARN deidentification system at learn.cybermango.org except in memory during the transitory moments when APPD LEARN ids are being generated.

5. Other

APPD LEARN studies do NOT have patients as participants, and thus never collect patient identifiers or health information.

Additional questions about APPD LEARN's de-identification system should be directed to:

Alan Schwartz, PhD, APPD LEARN Director Beth King, MPP, APPD LEARN Project Manager

Email: learn@appd.org

Sincerely yours,

Alan Schwartz, PhD, Director

Association of Pediatric Program Directors Longitudinal Educational Assessment Research Network (APPD LEARN)