Bioinformatics Integration Support Contract (BISC)

SUBJECT DE-IDENTIFICATION PROCESS (DRAFT)

Contract #: HHSN272200028C
Period Of Performance: September 30, 2012 to September 29, 2013
March 21, 2013

Project Sponsor:

National Institutes of Health (NIH)
National Institute of Allergy and Infectious Diseases (NIAID)
Division of Allergy, Immunology, and Transplantation (DAIT)

Prepared by:

NORTHROP GRUMMAN

Northrop Grumman Corporation
Information Systems Sector
Health IT
2101 Gaither Road, Suite 600
Rockville, MD 20850
(301)527-6400 Main
(301)527-6401 Fax
Jeff.Wiser@ngc.com

SUBJECT DE-IDENTIFICATION SUMMARY

The BISC team follows the Safe Harbor Method principles
(http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html) in
data capture templates and in data submission by:

- Rounding any "ages" submitted > 89 to 90 and add a comment indicating that this was done

- Not recording any dates within ImmPort templates; all time points and schedules are referenced
  to study day 0 (or a designated day in the study as determined by the data provider);

- Not recording any location information in ImmPort templates

- Not capturing any identifying information in ImmPort templates itemized in the Safe Harbor
  Method (Name, accounts, phone or fax numbers, email addresses, etc)

The BISC team severs the connection to subject identifiers in the source data management system by:

- If possible, surveying the data provider to determine whether local subject identifiers were used
  in the submission that would point back to a local system

- Hiding subject source identifiers, sample source identifiers, or experiment sample source
  identifiers in the ImmPort application to cut linkage with the source system

- Not exporting the subject source identifiers, sample source identifiers, or experiment sample
  source identifiers in any ImmPort generated data packages delivered online, at sFTP sites, or in
  the future, via Amazon AWS S3 storage

- Modifying result filenames to ImmPort accession numbers for individual result files (such as
  FCS, CEL, etc) if the filename may link back to subject source identifiers

- Not providing original, unstructured result files submitted by sites if they (1) haven't stated
  concurrence with the disconnect with source identifiers; and/or (2) if the data can be parsed into
  standard ImmPort templates and source identifiers hidden

The BISC team does not share GWAS level SNP genotyping data and Next Gen sequencing data within
ImmPort, and a formal request to NIAID similar to dbGAP is required for the data to be obtained

For data provided in more "native" format (i.e., not in ImmPort template format), such as result files from
closed clinical trials (assessments, lab tests, adverse events, etc):

- If the data source is ITN TrialShare, the dates are already obfuscated and are not identifying as
  are the subject identifiers; no special treatment is required

- For other data sources with closed clinical trial data (or the equivalent), data files are sent to our
  Stanford team's HIPAA server and evaluated by HIPAA trained staff to (1) Identify any
  potentially identifying information, such as names, dates, locations, etc... (2) Remove any
  identifying information and normalize dates to study day 0; (3) Comment and Description fields
  are removed, unless determined to be completely devoid of identifying information or identifiers;
  (4) Return the data to NG staff for submission and processing