1.0 POLICY

1.1 Background.

1.1.1 A common research method involves secondary analysis of publicly available survey data. The federal government provides public access to several important data sets (e.g., U.S. Bureau of the Census), and many federal funding programs now require that researchers make the data they collect publicly available. Likewise, many professional organizations and journals have a standard requirement that research data sets of published works be made accessible to encourage scholarly replication of research.

1.1.2 Under the federal regulations for human subjects research (45 CFR Part 46) publicly available data sets that are stripped of identifiers do not require IRB review. This policy describes the IRB’s determination about the circumstances under which the research use of publicly available data sets does not involve “human subjects”, and therefore does not require review and approval by the UA Institutional Review Board (IRB).

1.1.3 This policy does not deal with (1) restricted or limited use data sets (See POLICY: Limited Use Data Sets and POLICY: Restricted Use Data Sets), (2) merging public data sets in a manner that may increase the risk of identification of individual research participant; (3) Situations in which the data host requires the investigator or the investigator’s institution to sign a Data Use Agreement that explicitly requires IRB approval or a certification of exemption, or (4) research studies that also access non-public data and/or interaction or intervention with human subjects. These activities require IRB review or certificate of exemption.

1.1.4 Definitions

1.1.4.1 Public Data Sets are data files prepared by investigators or data suppliers with the intent of making data available for public use. The data are not individually identified or maintained in a readily identifiable form. The set has been reviewed by an IRB as appropriate for public use or, in the case of federal statistical collections, reviewed by the federal government.

1.1.4.1.1 De-identified means that the user of the data cannot readily ascertain the identity of the subject nor associate it with the information, and that the data set does not include any of the 18 HIPAA direct identifiers. If there is a code that links the data to direct identifiers, the code: (1) may not be derived from
or related to the information about the individual and (2) could not be translated to identify the individual.

1.1.4.1.2 **Public** means that the data are widely available. Data may still be considered public when a fee is charged for obtaining the data or when access is limited to researchers with academic or other research affiliations.

1.1.4.2 **Human Subject** is defined by federal agencies (DHHS, DoD, Veterans Affairs) as follow: A living individual about whom an investigator (whether professional or student) obtains (1) data through intervention or interaction with the individual or (2) identifiable private information. Intervention includes both physical procedures by which data are gathered (e.g., venipuncture) and manipulations of the subject or the subject’s environment that are performed for research purposes. Interaction includes communication or interpersonal contact between investigator and subject. Private information includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual, and which the individual can reasonably expect will not be made public (for example, a medical record). Private information must be individually identifiable (i.e., the identity of the subject is or may readily be ascertained by the investigator or associated with the information in order for obtaining the information to constitute research involving human subjects.

The **FDA definition** is “An individual who is or becomes a participant in research, either as a recipient of the test article [drug, device, or biological] or as a control. A subject may be either a health human or a patient. For device research, a subject is also an individual on whose specimen a device is used.”

### 1.2 Policy Statement.

1.2.1 It is the policy of the University of Alabama HRPP that the UA Research Compliance Office shall maintain a list of approved data sources for secondary analysis whose use does not require IRB approval so long as (1) the dataset meets the definition of a public data set, and (2) the investigator fulfills any requirements imposed by the owner/repository/source of the dataset.

1.2.2 The Research Compliance Office shall maintain and post on the website a Guidance document entitled “Public Data Sets Approved for Secondary Analysis Without IRB Review”.

1.2.3 Investigators may propose inclusion of new data sets on the list of approved data sets.

1.2.4 The IRB is the sole authority on whether or not research proposals using data sets require IRB review and approval.

### 1.3 Objective
1.3.1 Implementation of this policy assures investigators’ awareness of their roles and responsibilities in regard to IRB applications and protection of the privacy and confidentiality of human research participants and facilitates IRB protection of the same.

1.4 Responsibility

1.4.1 The Vice President for Research is ultimately responsible for this policy. Enabling parties include the Office for Research Compliance, the IRBs, principal investigators, and faculty supervisors of student research.

2.0 PROCEDURE

1.5 Investigator Use of Approved Data Sources

1.5.1 Investigators basing their research entirely within an approved data source (those identified on GUIDANCE: Public Data Sets Approved for Secondary Analysis Without IRB Review.

1.5.2 Investigators are welcome and encouraged to discuss their research plans with a Research Compliance Specialist or the Director of Research Compliance in advance of submitting their applications.

1.6 Investigator Submission of a Data Set for Pre-approval

1.6.1 Data sets that may qualify for inclusion on UA’s list of approved data sources include:

2.2.1.1 Public use data sets posted on the Internet that include a responsible use statement or other confidentiality agreement for authors to protect human subjects

2.2.1.2 Survey data distributed by UA principal investigators or external investigators who can attest that (1) the data collection procedures were approved by an IRB that satisfies the Common Rule criterion for an IRB, and (2) that the data set and documentation as distributed do not contain information that could be used to identify individual research participants.

1.6.2 To obtain pre-approved status for potentially eligible data sets, investigators should submit the following information to the Office for Research Compliance for review by a subcommittee of the IRB:

1.6.2.1 Name of data set;

1.6.2.2 URL of data set or other information on how to obtain the data set;

1.6.2.3 Abstract describing the content of the data set and its potential use;

1.6.2.4 IRB project approval number if the dataset was generated from research conducted at UA;
1.6.2.5 Name and other contact information of investigator in possession of the dataset if the dataset was generated from research conducted by an external investigator.

1.6.3 2.2.3 The Research Compliance Specialist will notify investigators if their data set has been granted pre-approved status and will update GUIDANCE: Public Data Sets Approved for Secondary Analysis Without IRB Review as needed.

2.0 REFERENCES

2.1 DHHS 45 CFR 46.102(f)

2.2 Department of Navy 32 CFR 219.102(f)

2.3 Department of Veterans Affairs 38 CFR 16.102

2.4 FDA 21 CFR 50.3(g), 21 CFR 812.3 (p).

2.5 National Human Subjects Protection Advisory Committee (NHRPAC): Recommendations on Public Use Data Files, January 28-29, 2002.

2.6 OHRP Guidance on Research Involving Coded Private Information or Biological Specimens, August 10, 2004

3.0 RELATED SECTIONS

3.1 GUIDANCE: Public Data Sets Approved for Secondary Analysis Without IRB Review

3.2 GUIDANCE: Template: Informed Consent for a Research Study

3.3 POLICY: Protection of Human Research Participants’ Privacy and Confidentiality

3.4 POLICY: Research Using Limited Data Sets

3.5 POLICY: Research Using Restricted Data Sets

3.6 GUIDANCE: The Meaning of Anonymous, Confidential, and De-Identified and Implications for Data Sharing or Re-use