SAMPLE DATA MANAGEMENT PLAN

(Submitted to the Social, Behavioral & Economic Sciences Directorate)

DATA MANAGEMENT PLAN

Expected Data

There are two main categories of data/information (all in digital format) that will result from this project:

- Curriculum materials in the form of software, content modules, assessments, and related information objects
- Qualitative and quantitative data and datasets collected or created as part of testing and ongoing assessments of the project and student learning; these data will be primarily in the form of text, numbers, and codes, and comprise the “raw” data for analysis; associated with the data are instruments, interview guides, protocols, analysis codebooks used to collect/capture/analyze project data, which will also be managed and available (in certain communities, some of these data objects are considered metadata, but for this data plan the term metadata relates to the library and information science context of this concept).

Period of Retention

Using facilities provided by the UNT Libraries’ Digital Library (http://digital.library.unt.edu), which is committed to long-term access and stewardship of publicly available research outputs, raw and derived datasets and information objects produced will be retained for at least 3 years after the end of the project, and may be available longer through UNT’s Digital Library. The project team will work with UNT Libraries in appraising the information objects for long-term preservation and to determine appropriate embargo periods for the individual digital resources. For data sources that are embargoed for some period of time, the metadata records will be available to allow discovery of the resources. All technical reports, presentations, and publications of the project will be available during the project through the web-based project site, and simultaneously or subsequently in UNT’s open access repository, the Scholarly Works Repository (http://digital.library.unt.edu/explore/collections/UNTSW), hosted in UNT Libraries’ Digital Library, with links from the project website to the resources in the Digital Library.

Data Formats and Dissemination

The project team will choose file formats for the digital data/information that will enable the most effective and most secure management, stewardship, and access over time. Formats for the textual data, when possible, will include, for example, PDFs, ASCII, CSV, and other well-known and open formats. Best practices for data curation and digital preservation will govern the choices of file formats all data/information. Where proprietary software is used, every effort will be made to provide the output of the software in as open a format as possible.

Metadata refers to descriptive, preservation, or other data about digital resources and will be used to describe and manage the project’s digital data/information. Metadata serves several purposes including: 1) description of the resource so a user will understand what the resource is, its context in the project, etc.; 2) discovery of the resource in a repository or digital library; 3) management and curation of the resource over time. The first two purposes can be generally achieved through the use of a descriptive metadata standard such as Dublin Core (http://dublincore.org/). This aligns with the standard metadata approach by UNT Libraries in its Digital Library, which will serve as one dissemination venue for publicly available project data and information products. For longer term management, curation, and preservation additional metadata (e.g., PREMIS metadata, http://www.loc.gov/standards/premis) will be used. This also aligns with UNT Libraries metadata and preservation standard practices.
Final versions of all curriculum materials (including software, content modules, assessment, etc.) will be deposited also at the Ethics CORE (Collaborative Online Resource Environment) Digital Library (http://nationalethicscenter.org/).

To the extent possible, data and derived datasets collected and/or created by the project will be available to other researchers for re-use. Decisions about open or limited availability will result from the appraisal discussed above. Data containing any personally identifying information will be anonymized prior to release for use to anyone outside of the project team. For data and datasets determined by the project team to require limited access, requests can be made to the Principal Investigator. Metadata describing all data and derived datasets will include information about requesting access to and use of the data and datasets (along with the associated codebooks, etc.). Others using the project’s collected, created, and/or derived data products will be requested to provide a standard citation (provided by the project) to credit both the NSF and this project for access and use of the data.

Data Storage and Preservation of Access

Data and information objects need to be managed during the project and stewarded for long-term access once the project has been completed. During the project, the College of Information through its Texas Center for Digital Knowledge will provide a web-based platform and virtual environment (built using Drupal, Islandora, and Fedora, http://islandora.ca/) to support project activities. Project team members will use this system for secure storage and sharing of all project data and resources. Access to this virtual research environment will be by authorized users only, using UNT’s authentication credentials. Authorized users will be able to access, analyze, and otherwise work with the data and information objects. All files on this web-based platform are covered by backup processes to ensure no unwarranted loss of project data. This online site provides the first step in data management and curation during the project.

At various points during the project, and especially at the end of the project, various project data and information objects will be submitted to UNT Libraries’ Digital Library. All technical reports, presentations, and publications of the project will be available during the project through the web-based project site, and simultaneously or subsequently in UNT’s open access repository, the Scholarly Works Repository hosted in UNT Libraries’ Digital Library, with links from the project website to the resources in the Digital Library. Near the end of the project and after appraisal of all project data/information products, the project team will work with UNT Libraries to ensure that all appropriate products are placed into the preservation environment provided by the Libraries for long-term stewardship and access. The UNT Libraries’ Digital Library incorporates preservation activities as part of storing and making accessible resources in the Digital Library. Long term access will also be assumed through the Ethics CORE Digital Library (http://nationalethicscenter.org/).
DATA MANAGEMENT PLAN

This data management plan (DMP) is intended to guide the secure storage and management of data gathered, created, and used during the project and indicates what and how project data will be shared, disseminated, and managed over time.

Types of Data

Both qualitative and quantitative data will be generated through various data collection activities during the project. These data will be primarily in the form of text, numbers, and codes, and these will comprise the “raw” data subject to analysis. The data may take the form of fieldnotes, recordings (audio and/or video) and transcripts of ethnographic fieldwork (e.g., interviews with participants, observations of behaviors), network data from ethnographic fieldwork, and electronic data from the virtual fieldwork. Summaries, and derived statistical reports, and visual depictions of select datasets presented as diagrammatic or mapped compositions will likely be generated. Instruments, interview guides, protocols, analysis codebooks used to collect/capture/analyze project data. In certain communities, some of these data objects are considered metadata, but for this data plan the term metadata relates to the library and information science context for this concept. In addition to created/collection data, the project may also collect secondary data source such as internal and/or archival Ecotrends documents.

Data and Metadata Standards

The project team will choose file formats for the digital data and information that will enable the most effective and most secure management over time. Formats for the textual data, when possible, will include PDF, ASCII, CSV, and other well-known and open formats. Best practices for data curation and digital preservation will govern the choices of file formats for text, audio, and video data. These formats reflect best practices for sharing, disseminating, and curating the data. To ensure the ability of others to make sense of, validate, and possibly reuse data created in the projects, datasets will be well documented in codebooks, which will include descriptions of how the data are structured, coding instructions, and how indices or scale scores were calculated. Where proprietary software is used, every effort will be made to provide the output of the software in an open format as possible.

Metadata refers to descriptive, preservation, or other data about digital resources and will be created to describe manage the project’s digital inputs and outputs. Metadata will serve several purposes including: 1) description of the resource so a user will understand what the resource is, its context in the project, etc.; 2) discovery of the resource in a repository or digital library; 3) management and curation of the resource over time. The first two purposes can be generally achieved through the use of a descriptive metadata standard such as Dublin Core (http://dublincore.org/). This aligns with the standard metadata approach by UNT Libraries in its digital library (http://digital.library.unt.edu/), which will serve as one dissemination venue for publicly available project results, outputs, data and information products. For longer term management, curation, and preservation additional metadata (e.g., PREMIS metadata, http://www.loc.gov/standards/premis/) will be used. This also aligns with UNT Libraries metadata and preservation standard practices.

Policies for Access and Sharing and Provisions for Appropriate Protection/Privacy

The Texas Center for Digital Knowledge at UNT will provide a web-based platform and virtual environment to support project activities. This will be used by project team members during the project for secure storage and sharing of all project data and resources. This will be the platform for the case study database described in the Research Design section of the proposal. Access
to this virtual research environment (built using Drupal, Islandora, and Fedora, http://islandora.ca/), will be by authorized users only and who must use UNT’s authentication credentials. Use of this system will allow access by the authorized individuals to access, analyze, and otherwise work with the data. All files on this web-based platform are covered by backup processes to ensure no unwarranted loss of project data. This online site provides the first step in data management and curation during the project. It will also serve as the initial discovery path for project created resources.

Team members will appraise the source and derived data and datasets created during the process to determine potential benefits to future researchers, and will make those available publicly through the UNT Libraries’ Digital Library. Data containing any personally identifying information will be anonymized prior to release for use to anyone outside of the project team. The project team will work with UNT Libraries to determine appropriate embargo periods for the individual digital resources. For data sources that are embargoed for some period of time, the metadata records will be available to allow discovery of the resources. All technical reports, presentations, and publications of the project will be available during the project through the web-based project site, and simultaneously or subsequently in UNT’s open access repository, the Scholarly Works Repository (http://digital.library.unt.edu/explore/collections/UNTSW/), hosted in UNT Libraries’ Digital Library, with links from the project website to the resources in the Digital Library.

Policies and Provisions for Re-Use, Re-Distribution
To the extent possible, data and derived datasets collected and/or created by the project will be available to other researchers for re-use. Decisions about open or limited availability will result from the appraisal discussed above. For data and datasets determined by the project team to require limited access, requests can be made to the Principal Investigator. Metadata describing all data and derived datasets will include information about requesting access to and use of the data and datasets (along with the associated codebooks, etc.). Others using the project’s collected, created, and/or derived data products will be requested to provide a standard citation (provided by the project) to credit both the NSF and this project for access and use of the data.

Plan for Archiving, Preservation, and Access
All data stored in the virtual research environment (in the Fedora repository) is routinely backed up and serves as a first layer for curating and archiving during the project. The project’s online site will be maintained for three years following the project end date. Near the end of the project and after appraisal of all project data products, the project team will work with UNT Libraries to ensure that all appropriate project products are placed into the preservation environment provided by the Libraries for long-term stewardship. The UNT Libraries’ Digital Library incorporates preservation activities as part of storing and making accessible resources in the Digital Library.
Data Management Plan
for NSF ADVANCE Proposal

This data management plan (DMP) is intended to guide the management of data gathered, created, and used during the project and indicate what and how project data will be shared, disseminated, and managed over time.

Types of Data: Both qualitative and quantitative data will be generated through various data collection activities during the project. These data will be primarily in the form of text, numbers, and codes, and these will comprise the "raw" data subject to analysis. Some data may be captured as audio (e.g., interviews with participants). Summaries and derived statistical reports will be generated through analysis. Data collection instruments and protocols used to capture data will be developed for the project and/or taken from the public domain. Other resources produced in the project will be codebooks associated with the created datasets. The project will also collect a wide range of information and data as inputs into the study activities. These may include presentations, articles, chapters, policies, and procedures from other universities. Internal UNT documents, such as reports by consultants hired by UNT, may be used in the project. Where these resources are in digital format, they will be considered for dissemination and access.

Data and Metadata Standards: The project team will choose file formats for the digital data and information (both inputs to and created during the project) that will enable the most effective management over time. Formats for the data, when possible, will include PDFa, ASCII, CSV, and other well-known and open formats. These formats reflect best practices for sharing, disseminating, and curating the data. To ensure the ability of others to make sense of and possibly reuse data created in the project, datasets, especially numerical datasets, will be well documented in codebooks, which will include a description of how the data are structured, recoding instructions, and how indices or scale scores were calculated. For example, a copy of the instrument for the social science study will be made with labels for the variables, citations for public domain scales with the specific items identified, recodes and computed variables inserted.

Metadata records will be created to describe each of the project digital resources (e.g., datasets from surveys, transcripts from interviews, etc.). Metadata will serve several purposes including: 1) description of the resource so a user will understand what the resource is; 2) discovery of the resource in a repository or digital library; 3) management and curation of the resource over time. The first two purposes can be generally achieved through the use of a descriptive metadata standard such as Dublin Core (http://dublincore.org/). This aligns with the standard metadata approach by UNT Libraries in its digital library (http://digital.library.unt.edu/), which will serve as one dissemination venue for publicly available project results, outputs, data and information products. For longer term management, curation, and preservation, additional metadata (e.g., PREMIS metadata, http://www.loc.gov/standards/premis/) will be used. This also aligns with UNT Libraries metadata and preservation standard practices.

Policies for Access and Sharing and Provisions for Appropriate Protection/Privacy: UNT uses Sharepoint as its intranet system, and this will be used by project team members during
the project for secure storage and sharing of project data and resources. Access to Sharepoint is by authorized users only, who must use UNT's login and password credentials. Use of this system will allow access by the individuals allowed to access, analyze, and otherwise work with the data. All files on Sharepoint are under centralized backup processes, to ensure no unwarranted loss of project data.

The project team will have access to data generated and/or held in various UNT offices. Such data may include salary, stipends, rank, merit evaluations, resources, promotion information, and search committee details. Because UNT is a public institution, strictly speaking, these data are not confidential, since all such information is available for review. However, this project will replace actual names in the data with unique codes or numbers (de-identify). The Senior Internal Evaluator (IE) will create a list that records unique number assigned to each UNT faculty member, staff member, and administrator whose name may appear in the raw data. Only the internal evaluators, Principal Investigator (PI) and Project Director (PD) will have access to this list. This list will not be kept on the intranet but in a secure physical location (e.g., a locked file cabinet in the PI's office).

Project team members will appraise the source data and datasets created during the process to determine potential benefits to future researchers, and will make those available publicly through the UNT Libraries Digital Library. In some cases, there may need to be some aggregation of data (e.g., a department with one woman may have to be collapsed with a similar department that has several women or data may have to be collapsed for type of discipline such as life sciences). The PD and IE will determine if data should be collapsed before release. The project team will work with UNT Libraries to determine appropriate embargo periods for the individual digital resources. For data sources that are embargoed for some period of time, the metadata records will be available to allow discovery and description of the resources. All reports, presentations and publications of the project will be available through the UNT Digital Library, with links from the project website to the resources in the Digital Library.

Policies and Provisions for Re-Use, Re-Distribution: To the extent possible, data and datasets created by the project will be available to other researchers for re-use. Public availability decisions will result from the appraisal discussed above. For data and datasets determined by the project team to have limited access, requests can be made to the Project Director (and to the Vice Provost for Faculty Development after the end of funding). Metadata describing all data and datasets will include information about requesting access to and use of the data and datasets (along with the associated codebooks, etc.).

Plan for Archiving and Preservation of Access: Although UNT's Sharepoint is not a preservation mechanism, all data stored on the system is routinely backed up and serves as a first layer for archiving during the project. The UNT Libraries' Digital Library (DL) incorporates preservation activities as part of storing and making accessible resources in the DL. The DL also hosts UNT's open access institutional repository where the project's data, reports, and other products can be stored (http://digital.library.unt.edu/explore/collections/UNTS/WI). The project team will work with UNT Libraries to ensure that all appropriate project products are placed into the preservation environment provided by the Libraries.