



# Safe to Save? Archive Options for Geodatabases

Jeff Essic  
North Carolina State University Libraries

## GeoMAPP

- **G**eospatial **M**ultistate **A**rchive and **P**reservation Partnership
- Partnership between state archives, GIS data centers, and libraries under the National Digital Information Infrastructure and Preservation Program (NDIIPP)
- Focus on state and local geospatial content; Long-term preservation (10, 20, 50, ... years)
- Partnership Website: <http://www.geomapp.net>
- NCSU Website: <http://www.lib.ncsu.edu/ncgdap>



## Final Report

### **“Archival Challenges Associated with the ESRI Personal Geodatabase and File Geodatabase Formats”**

Planned to be released in Fall, 2011

Watch <http://www.geomapp.net/>

# Geospatial Data Preservation Challenge: Vector Data Formats

- No widely-supported, open vector formats for geospatial data
  - Spatial Data Transfer Standard (SDTS) not widely supported
  - Geography Markup Language (GML) – diversity of application schemas and profiles a challenge for “permanent access”
  
- Spatial Databases
  - Valuable components: relationships, rules, topologies, domains, networks, linked annotations, metadata, etc.
  - The whole is more than the sum of the parts, but the whole is very difficult to preserve

## Geodatabase Curation Study: Overview

- Three types of ESRI Geodatabases (GDB):
  - SDE, Personal, File
- Practical curation/conversion options:
  - Archive the GDB object
  - Export to: XML, shapefiles, GML Simple Features (open published formats)
- **Question:** Will GDB objects and export files created in older ArcGIS versions be compatible with newer versions?

## Constraints

- Only tested the most reasonable and logical conversion scenarios for an archivist, and those with the highest potential for failure. Numerous other possibilities not tested.
- Didn't explore GDB's with rasters.
- Very limited sample sizes and test executions

## SDE Geodatabase

- Commonly used among local governments for enterprise data management
- Stored in RDBMS, so can't be archived as a stand-alone "snapshot" object.
- Must be exported to XML, fGDB, and/or shapefiles
- Supports Historical Archiving

## Personal Geodatabase

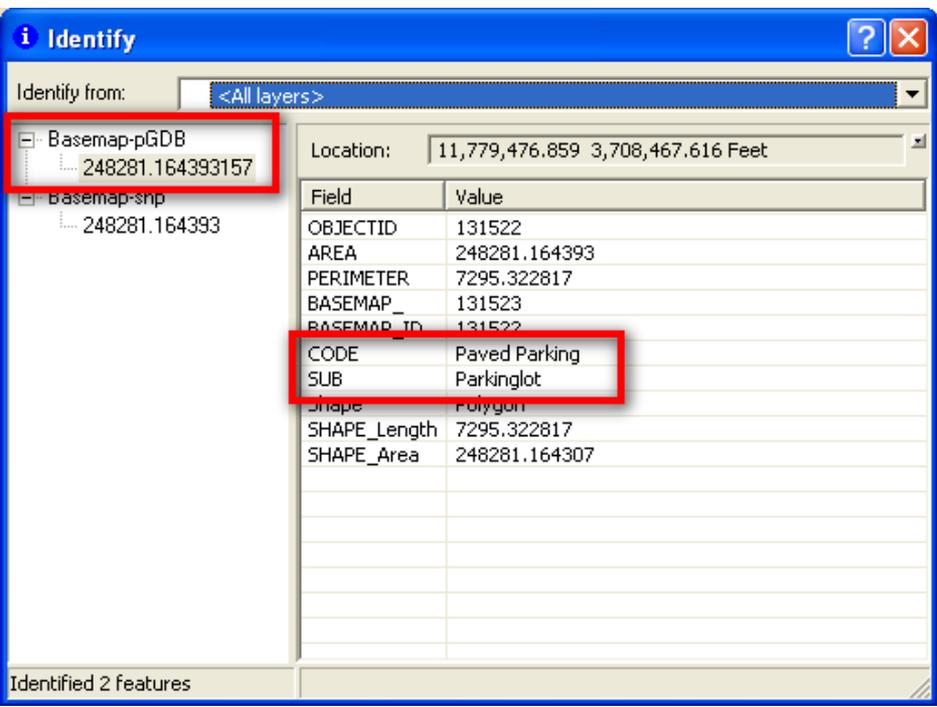
### ■ Not ideal archival object

- Very proprietary – ArcGIS / MS Access formats
- ESRI recommends using File GDB instead

[http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Types\\_of\\_geodatabases](http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Types_of_geodatabases)

### ■ Archive export formats: XML, fGDB, and/or shapefiles

# pGDB Export to Shapefiles



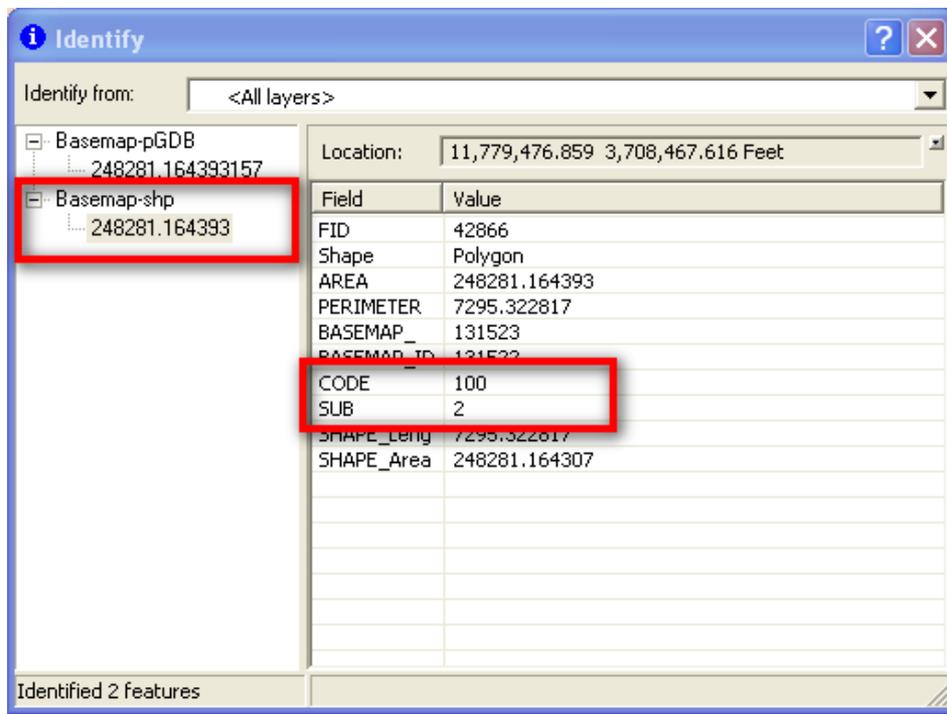
Identify from: <All layers>

- Basemap-pGDB
  - 248281.164393157
- Basemap-shp
  - 248281.164393

Location: 11,779,476.859 3,708,467.616 Feet

Field	Value
OBJECTID	131522
AREA	248281.164393
PERIMETER	7295.322817
BASEMAP_	131523
BASEMAP_ID	131522
CODE	Paved Parking
SUB	Parkinglot
Shape	Polygon
SHAPE_Length	7295.322817
SHAPE_Area	248281.164307

Identified 2 features



Identify from: <All layers>

- Basemap-pGDB
  - 248281.164393157
- Basemap-shp
  - 248281.164393

Location: 11,779,476.859 3,708,467.616 Feet

Field	Value
FID	42866
Shape	Polygon
AREA	248281.164393
PERIMETER	7295.322817
BASEMAP_	131523
BASEMAP_ID	131522
CODE	100
SUB	2
SHAPE_Leng	7295.322817
SHAPE_Area	248281.164307

Identified 2 features

Sub-domain attribute text is lost in the conversion to shapefile

## File Geodatabase

- Preferred stand-alone ESRI format for large dataset storage – numerous usability advantages
- Growing potential as archival object
  - Kentucky KYGEONET
- Archive export formats: XML, shapefiles, GML



## File Geodatabase

### KYGEONET:

**“Snapshot File Format – Kentucky has chosen to archive its data in the form of an ESRI’s file-based geodatabase (fGDB).** This file-based relational database format will allow the entire archive set to exist within its own container with groupings of data based upon the FGDC Metadata model (same as groupings on KYGEONET and GOS). This file format is appropriate for the storage of both raster and vector data and allows for compression. Additionally, the fGDB allows for vector topology, the inclusions of route data, and other advanced relationships that cannot be supported with the old Shapefile format.”

[http://www.geomapp.net/docs/ky\\_geoarchives\\_procedures.pdf](http://www.geomapp.net/docs/ky_geoarchives_procedures.pdf)

## File Geodatabase

### ■ Advantages:

- File Geodatabase API – potential for development of third-party read/write/access tools
- Folder/File structure: can see “under the hood”
- High level of ESRI support

### ■ Disadvantages:

- fGDB API has limitations re: dataset types and rasters
- No fGDB open specification
- Metadata is not externally accessible
- Version issues: older fGDBs require upgrading

## File Geodatabase API

- Opens fGDB access to a limited extent to non-Esri applications
- Potential for broader fGDB support, including by non-Esri software, helping increase the likelihood of longer-term usability and transparency of data and increased sharing of fGDB formatted data
- Limitations: support of various dataset types, raster components, and spatial query operations

## Conversion Options – Entire Geodatabase

### ■ Geodatabase Upgrade

- Regularly upgrade geodatabases to new version
- Labor intensive

### ■ ArcGIS XML

- Exports all items and data within GDB
- Early standard -- not widely adopted and tested
- XML files are quite large. Can be compressed, but that's another potential point of failure

## Conversion Options – Individual Datasets

### ■ Shapefile

- Open spec, very stable, widely accepted
- No relationships and other GDB components export
- Limits on column names, file sizes

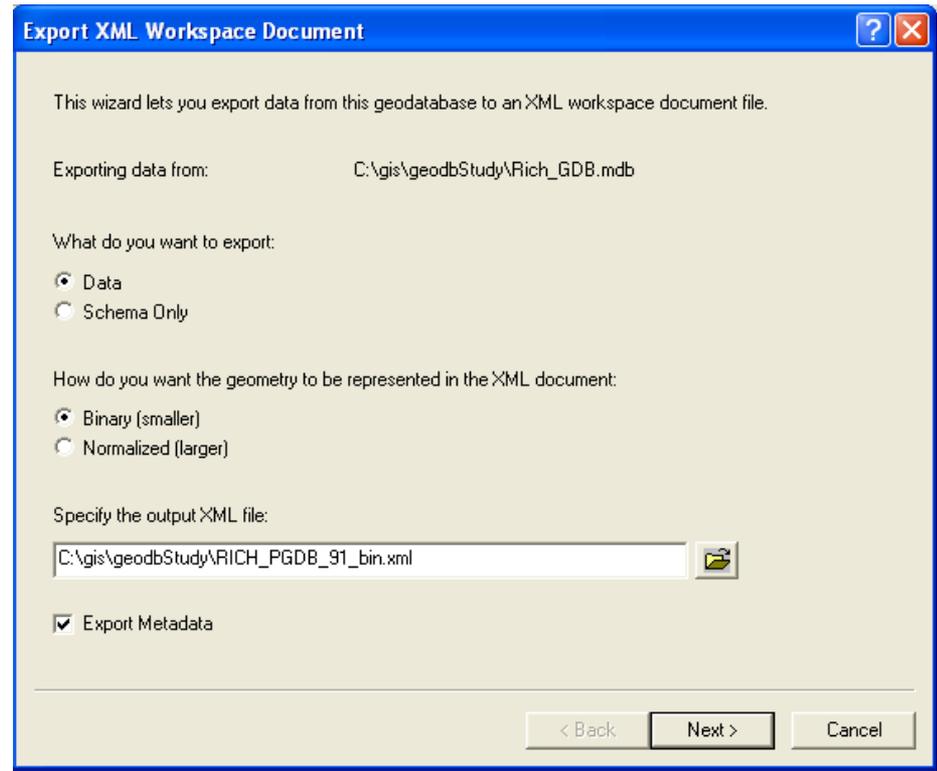
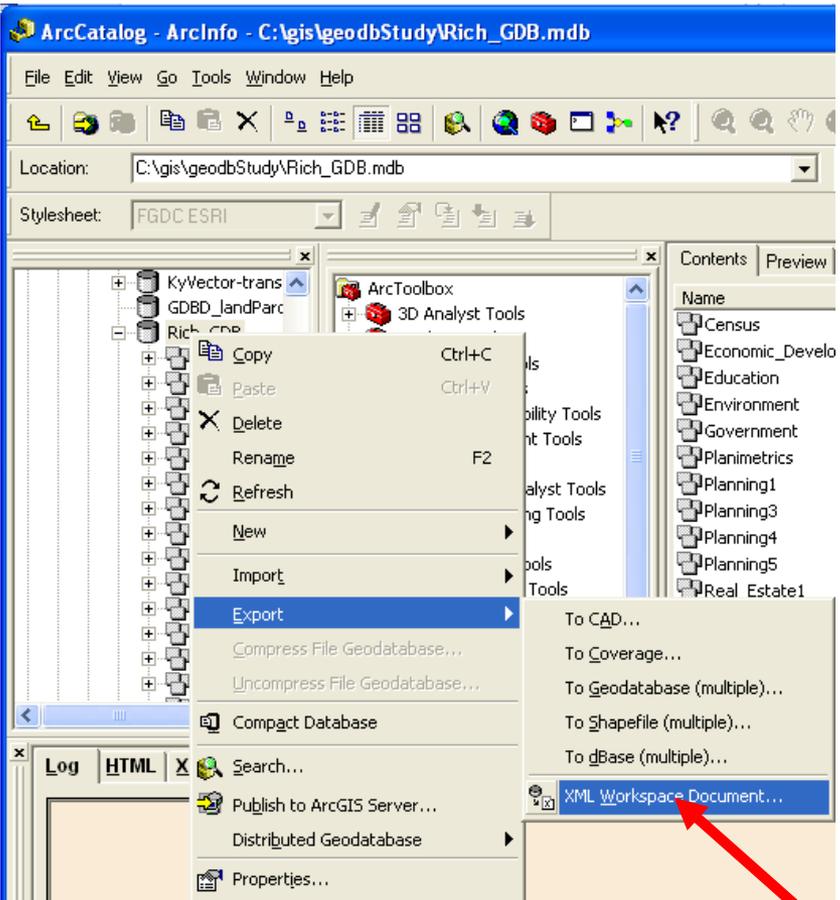
### ■ Simple Feature GML

- Open standard of OGC
- Numerous GML flavors – how they fit within an archive context is questionable
- No current batch conversion mode

## Highlights of Earlier Testing

- Previously tested import/export (“round-tripping”) older geodatabases using ArcGIS 9.3.1
- Results: [http://www.lib.ncsu.edu/ncgdap/presentations/esri09\\_essic.ppt](http://www.lib.ncsu.edu/ncgdap/presentations/esri09_essic.ppt)
- Highlights:
  - 8.3 pGDB export to 9.1 XML was successful, but importing this XML to all newer pGDB versions consistently failed.
  - Upgrading the pGDB first, and then converting to XML was successful.
  - Export of a 9.2 fGDB to XML, then import this XML to a 9.3.1 fGDB was successful.
  - Export of a 9.2 fGDB to GML, then import this GML to a 9.3.1 fGDB initially failed, but later testing was successful.

# Personal and File GDB Export



Export to XML interface

Export to XML

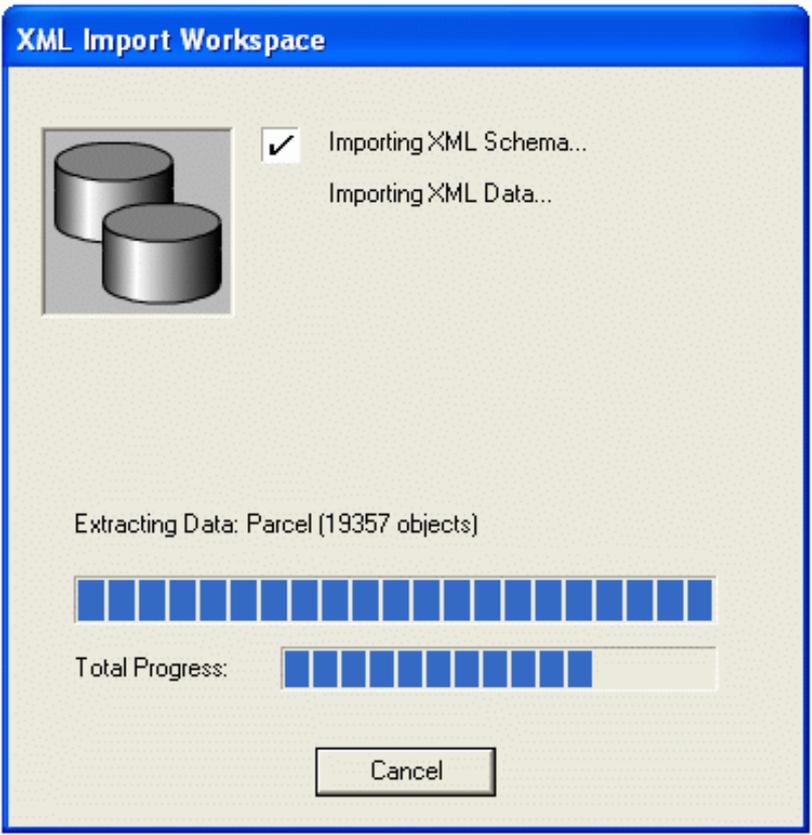
# Personal GDB Tests

Richmond VA pGDB – Version 8.3 – Created October 3, 2003

	Initial Size	Compressed Size	Ratio
Original 8.3 pGDB	728 MB	309 MB	1:2.36
Export to 9.1 XML using Binary	<b>Success</b> 2.8 GB (4X > than source)	269 MB	1:10.7
9.1 XML Import to pGDB using 9.1	<b>Success</b> 736 MB <b>Attribute text for Sub-Domains and Relationships Preserved</b>		
9.1 XML Import to pGDB using 9.2	<b>FAILED</b> (size reached 394 MB)		
9.1 XML Import to pGDB using 9.3.1	<b>FAILED</b> (size reached 788 MB)		
9.1 XML Import to pGDB using 10.0	<b>FAILED</b> (size reached 411 MB)		

These tests measure the reliability of importing an XML of a pGDB several years/versions after it was created.

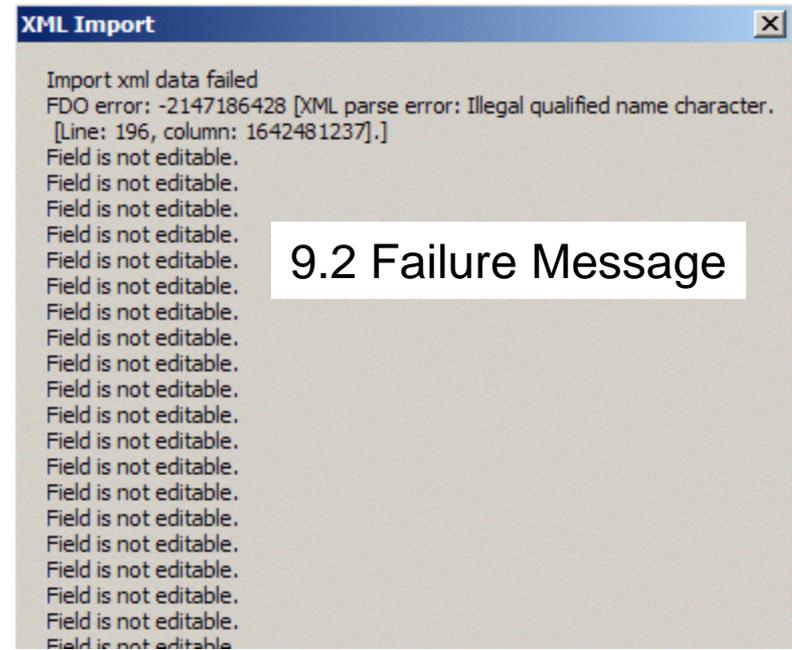
# pGDB Import of 9.1 XML



Import in progress



9.3.1 Failure Message



9.2 Failure Message

# pGDB Upgrade to 10.0

Richmond VA pGDB – Version 8.3 – Created October 3, 2003

	Initial Size	Compressed Size	Ratio
Original 8.3 pGDB	728 MB	309 MB	1:2.36
Upgraded to 10.0 pGDB	<b>Success</b> 730 MB Note: Upgrade using “Properties/Upgrade Geodatabase”		
Export to 10.0 XML	<b>Success</b> 1.25 GB		
XML Import to 10.0 pGDB	<b>Success</b> 738 MB Functionality and content intact		

These tests measure the reliability of XML export / import “round-tripping”

# pGDB conversion to fGDB

Richmond VA pGDB – Version 8.3 – Created October 3, 2003

	Initial Size	Compressed Size	Ratio
Original 8.3 pGDB	728 MB	309 MB	1:2.36
Upgrade to 10.0 fGDB	<b>Success</b> 225 MB / 297 Files File size and number of files is less than 9.3.1 upgrade (274 MB/322 files)		

This tests the reliability of upgrading an old pGDB to a present version fGDB.

# File GDB Tests

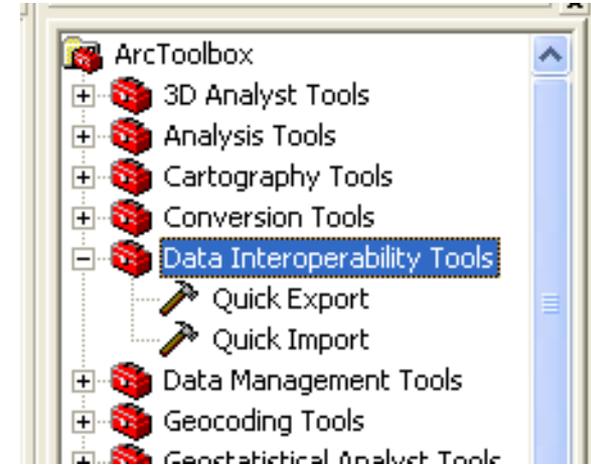
Kentucky Transportation Vectors – Version 9.2 – Acquired 6 June 2009

	Initial Size	Compressed Size	Ratio
Original 9.2 fGDB	224 MB / 64 files	80.9 MB	1:2.77
Export to 9.2 XML using Binary	<b>Success</b> 1.11 GB (5X > than source)	137 MB	1:8.3
9.2 XML Import to 10.0 fGDB	<b>Success</b> 224 MB / 77 Files		

This test measures the reliability of importing an old XML of a fGDB to a present version fGDB.

## GML Export

- ArcToolbox/Data Interoperability Tools:  
GML “Simple Features Profile”  
support available out-of-the-box to  
all users
- 9.2 fGDB → 9.3.1 GML → 9.3.1 fGDB:  
First conversion test failed, but later re-test and all other  
later tests were successful
- Conversions are very laborious -- each feature class  
must be selected



## Testing Summary

### ■ Failures:

9.1 XML import

9.3.1 GML import (later successful)

■ May be due to early bugs that have been corrected, but a slight lack of confidence remains.

■ All other conversion and round-trip tests were successful



## Geodatabase Preservation Strategy

- Based on Digital Preservation Coalition Technology Watch Report “Preserving Geospatial Data” 5/2009

[http://www.dpconline.org/component/docman/doc\\_download/363-preserving-geospatial-data-by-guy-mcgarva-steve-morris-and-gred-greg-janee](http://www.dpconline.org/component/docman/doc_download/363-preserving-geospatial-data-by-guy-mcgarva-steve-morris-and-gred-greg-janee)

- Multi-prong approach
- Assumes short-term format support from Esri and Safe FME products
- Long-term support will always be an unknown
- Key is to avoid neglecting the data

## Geodatabase Preservation Strategy

- ArcSDE → maintain temporal versions, and export fGDB and possibly shapefiles in a separate archive
- fGDB → maintain as fGDB. Optionally, convert feature classes to shapefiles in addition to retaining fGDB
- Shapefiles → retain in shapefile format, try to acquire original GDB
- pGDB, XML, GML, or other format → Convert to fGDB to reveal any potential conversion errors
- Upgrade fGDB content to newer versions as they become established

## What's Next?

- Finalize Report
- Future study: SDE Export Options – Best practices to preserve as much as possible via XML, fGDB, and/or shapefiles?
- Keep an eye on open source options (fGDB API implementation, SpatiaLite)
- Provide improvement suggestions to Esri



## *Slide Presentation*

<http://www.geomapp.net/presentations.htm>

Jeff Essic  
Data Services Librarian  
NCSU Libraries  
[jeff\\_essic@ncsu.edu](mailto:jeff_essic@ncsu.edu)