SECTION II: Data Collection and Methods of Analysis

II. A - Overview

This section provides an overview of the data sources and software used in the creation and maintenance of data for the study, a discussion of existing data collected from various providers, a description of the University's data creation methodology, and a brief outline of the resulting database structure.

The University spent 15 months collecting and analyzing data. The data collection process proved cyclical and continuous, continuing into the spring of 2010. Much effort during the study was spent collecting available data and transforming and combining these items into user friendly products for analysis. Important gaps in the data were supplemented with University field studies.

The study area for this report included ten counties in Pennsylvania where active underground bituminous coal mining occurred during the study period. They were: Armstrong, Beaver, Cambria, Clearfield, Elk, Greene, Indiana, Jefferson, Somerset, and Washington. See Section III for a breakdown of mining in each county.

The University amassed a large spatial and non-spatial database for the study area. The deliverables required the creation of tabular data, spatial data, photographic materials, and the collection of biological specimens. The resulting datasets of tabular and spatial data were approximately 900-MB and 53-GB, respectively. Tabular data, about features on the surface, and spatial data, enabling identification of spatial relationships between features, were combined to generate a powerful tool for analyzing the effects of underground bituminous coal mining on surface features.

II.B - Data Sources and Software

Spatial and non-spatial data were gathered to answer the questions posed in the scope of work. The data gathering effort spanned 15 months and continued until the end of the contract. Spatial data, in this study, refers to records in a spatial format which includes both location (coordinate) and attribute (descriptive) information. Non-spatial data is used to refer to records that do not include location information such as spreadsheets, photographs, paper documents, etc. Sources of spatial data include:

- PASDA geospatial clearinghouse,
- PA DEP CDMO and the Bureau of Mining and Reclamation (BMR) in Harrisburg, PA,
- The U.S. Census Bureau,
- The U.S. Geological Survey (USGS),
- The Western Pennsylvania Conservancy,
- Several mining operators voluntary submissions of digitized maps, and
- University generated data based on field research.

Formats for spatial data include Shapefiles, Geo-database files, Interchange files, Geo-Tagged Image Format (Geo-TIF) files, and Computer Aided Drawing (CAD) files. Shapefiles, and Geo-database files are proprietary formats used by ESRI's ArcEditor application, while Interchange files are designed for compatibility between Geographic Information System (GIS) applications. Geo-TIF files are similar to Tagged-Image Format (TIF) files with the exception that the Geo-TIF includes location information. Finally, CAD drawing files are proprietary formats used by AutoDesk's CAD application.

Sources for spatial data include PA DEP CDMO, PA Department of Conservation and Natural Resources (PA DCNR), and several mining operators. Sources for non-spatial data include Microsoft Excel spreadsheets, Joint Photographic Experts Group (JPEG) images, TIF files, and paper files.

The accumulation of digital data necessitates the use of applications to read and manipulate the information. As a result, various software programs are employed for data collection and analysis. Table II-1 lists the software programs used in this study, their use, and year of publication.

Table II-1 – Software programs used in this study, use, and year of publication.

Programs	Use	Year
Adobe Photoshop *CS4*	Image Processing Software	2008
ArcGIS Desktop Evaluation Edition *v.9.3*	Mapping and spatial analysis	2008
Analysis of Retreat Mining Pillar Stability (ARMPS)	Pillar Stability	2009
AutoCAD	View and export data	2007
Garmin MapSource	GPS Data Importation	2005
Irfanview	Image Processing Software	2008
Microsoft Office Suite	Data Tabulation and Analysis	2003 & 2007
Microsoft Vista or XP	Operating System	2006
Surface Deformation Prediction Software	Subsidence Prediction	2006

II.C – Collection of Existing Data

Building a database with the capability to analyze the surface impacts of underground bituminous coal mining began with the acquisition of various types of existing data. The tasks set forth in the scope of work merited the collection of base data such as roads and streams, 6-month mining maps, and non-spatial information regarding surface features undermined. The University also obtained access to PA aquatic community data, and a collection of AutoCAD maps that included areas and features undermined. This section details the collection and manipulation of existing data collected during the study. The data described below provided the building blocks for the subsequent tasks of data creation and spatial analysis.

II.C.1 – Base Data

In this study, base data is defined as spatial data with locations of transportation networks, hydrologic features, and political boundaries. These kinds of spatial layers were acquired from the PASDA geospatial clearinghouse maintained by The Pennsylvania State University (PSU) and the Geography Division of the U.S. Census Bureau. They include the following:

- Road datasets:
  - Local and state road datasets created by the PA Department of Transportation (PennDOT), and
  - Topologically Integrated Geographic Encoding and Referencing System of the U.S. Census Bureau.
- Hydrologic datasets:
  - Small watersheds generated from the USGS Water Resources Division's major watersheds dataset by the Environmental Resources Research Institute (ERRI) at PSU,
  - Waterbodies from the National Hydrography Dataset (NHD) created by the USGS,
  - Networked streams of PA created by ERRI,
  - PA Chapter 93 streams with designated use created from the NHD by PSU Institutes of the Environment and Research Triangle Institute, and
  - PA DEP CDMO's GIS/Global Positioning System (GPS) stream observation database.
- Political boundaries datasets:
  - Statewide PA county boundaries created by PennDOT, and
  - Individual county outlines created by CUP during the 2nd assessment.

II.C.2 – Aquatic Community Data

The Aquatic Community Classification Database (ACCD) is both a project and product of the PA Natural Heritage Program of the PA DCNR. The project's goal is to describe communities of fish, mussels, and macroinvertebrates found in streams within PA. The ACCD includes tabular and spatial datasets detailing these aquatic communities, associated geology, stream reaches and watershed boundaries. This database proved useful for analysis related to macroinvertebrate populations and water supply impact classifications. The University obtained access to this database through the Western Pennsylvania Conservancy.

II.C.3 – 6-Month Mining Maps

One core dataset was the collection and consolidation of 6-month mining maps for each mine operating during the study period. These maps were obtained from the PA DEP CDMO. As required by Department regulations, operators submit mine maps at a predetermined scale to the CDMO every six months. This is the PA DEP method for monitoring the progression of mining. The maps are called 6-month mining maps because they depict the location of the last six months of mining prior to the creation date, and a forecast of projected mining during the impending six months. The maps also include information regarding the locations of surface features such as properties, structures, water supplies and utilities. Some maps include coal and surface elevation contours and road and stream networks. An estimated 900 6-month mining maps were received as TIF images by the University and after duplicate and inactive mine maps were eliminated, 317 were used in the assessment.

II.C.4 – Bituminous Underground Mining Information System (BUMIS) Inventory

The PA DEP uses a database called the BUMIS to maintain an inventory of surface features related to underground bituminous mining activities. Surface features are defined as properties, structures, water supplies, utilities, and streams. The University queried the BUMIS database for surface features undermined during the 3rd assessment period, as well as any reported effects. For an explanation of inventory methods for undermined streams refer to Section VIII.

The initial search was conducted by querying each mine name to discover the associated properties undermined between the dates of August 21, 2003 and August 20, 2008, and their corresponding BUMIS System ID number. The second search utilized the BUMIS System ID number for undermined properties to harvest the associated structures, water supplies and utilities undermined for each property. The collected data was organized, sorted, and maintained using Microsoft Excel spreadsheets. Tables were produced for each mine in the study area.

The property numbers in the tables were then compared with property numbers on the 6-month mining maps. Sub-section II.E.3.a has a more detailed discussion of the property rectification procedures. After comparison, any new properties found were queried in BUMIS, and, if found, were added to the University's features undermined inventory. Additionally, several properties were removed from the inventory because they were found to be outside the University's criteria for the active mining region.

After all available information was collected from BUMIS; the University generated final tables for each mine. A Mine Information Spreadsheet and a Final GIS Spreadsheet for each mine were created. These were essentially the summation of the spatial data in tabular form.

The purpose of the Mine Information Spreadsheet is to display the BUMIS inventory and more detailed comment information for features with reported effects. The purpose of the Final GIS Spreadsheet is to enable the collected BUMIS information to join with the spatial attribute table for each feature type in each mine. Sub-section II.E.3 discusses the record matching process for appending the BUMIS information with University generated spatial data.

II.C.5 – Operator Data

University field representatives visited several operators' offices to obtain more accurate digital spatial data regarding mining outlines, surface features undermined, and digital tabular data for Module 8: Hydrologic Information of the permit files. The CDMO has traditionally obtained these materials in paper format, and while this process is migrating toward digital format, during the 3rd assessment period the method of collection was still printed form. Therefore this data had to be manually entered into the electronic repositories.

II.C.6 – Elevation Data

Surface elevation layers were obtained from the National Elevation Dataset available from the USGS. Two digital elevation models (DEM) were acquired for the study area, one covering

Beaver, Washington and Greene counties and a second covering Armstrong, Cambria, Clearfield, Elk, Indiana, Jefferson, and Somerset counties. The spatial resolution upon acquisition was 30-meters (98.4-ft).

Coal contour and some overburden data were collected from Chapter C and Chapter D of "Resource Assessment of Selected Coalbeds and Zones in the Northern and Central Appalachian Basin Coal Regions" (Anon, 2000). These contours proved too coarse to use for overburden generation over each mine, but provided a useful dataset that encompassed the entire study area and allowed comparison between USGS and University generated overburden values.

II.C.7 – Pennsylvania Groundwater Information System

The PA DCNR's Topographic and Geologic Survey compiles and maintains an online database for groundwater information called the PA Groundwater Information System (PaGWIS). Tabular data for groundwater in the study area was obtained from this online resource and imported into a spatial dataset using x, y, coordinate values.

II.C.8 – ACT 54 2nd Assessment Spatial Data

The University acquired spatial data utilized during the 2nd assessment report at the initial stages of this study. That dataset was created by CUP (as a course of completing the 2nd assessment) and included geo-referenced 6-month mining maps, and spatial data for most longwall mines and a small number of room-and-pillar mines. This data was generally used for comparison with several datasets for longwall mines created during this assessment.

II.D – Data Standardization

Each spatial dataset needed a coordinate system to display its location on the Earth's surface. This coordinate system measured the respective locations using geodetic coordinates such as latitude and longitude, Cartesian grid coordinates, or polar coordinates. Once all existing spatial datasets were collected, they were re-projected to the Universal Transverse Mercator (UTM) Coordinate System, using the North American Datum of 1983 and the map projection for UTM Zone 17 North.

The use of a map projection allowed areas and distances to be calculated. A map projection (a mathematical technique) was used to translate the spherical dimensions of the earth (3D) to the planar dimensions of a flat map (2D) (DeMers, 2003). The combination of a coordinate system and a map projection is called a spatial reference.

Several of these datasets required geographic or coordinate transformation because they used a different datum. A datum is a reference surface used to model the shape or size of the Earth (DeMers 2003). Datum are used as the reference surface in most coordinate systems. To transform the spatial reference of a dataset to the reference standard, a coordinate transformation was performed. Coordinate transformation uses one of several mathematical models to adapt the input coordinate system to match the desired format.

II.D.1 – Spatial Reference

The North American Datum of 1983 was chosen as the reference datum because it is based on a newer spheroid, the Geodetic Reference System of 1980. While the North American Datum of 1927, (based on the Clarke spheroid from 1866) was still being used by some entities, and by older datasets, the use of the newer earth shape model for this study was warranted.

The UTM system is used for the entire earth, whereas, the state plane system is exclusive to the US. The UTM system is composed of 60 zones each equaling 6-deg in width. The zones are oriented in a north-south direction; therefore, the state of Pennsylvania is cut through the center leaving an eastern and western half. All mines in the study area are encompassed by zone 17 within the UTM system. This knowledge, coupled with the use of the UTM projection for data in the 2nd assessment led the University to choose the UTM projection as the reference standard for the study.

Acquired base data were produced using various map projections and either NAD 1983 or NAD 1927. These data were re-projected to the reference standard used in the study. All 6-month mining maps, digitized features and PA GWIS point dataset were assigned this reference standard during the creation of the dataset.

II.D.2 – Use of the Geo-Database Format

Base and PA Aquatic Community Data were imported to and stored in the geo-database. ESRI's ArcGIS personal geo-database does not allow simultaneous editing, however, multiple users can view contents and the database can hold up to 2-GB of spatial and tabular data. A geo-database of this kind was populated for each mine in the study. Data within each mine geo-database was trimmed to the extent of the respective county of operation. Geo-databases produced using these methods were incorporated in the final database structure.

II.E –Data Creation

As noted in Section II.C, a considerable amount of data was collected in the initial stages of the study. Most of these items were used to generate necessary data to answer the questions posed in the contract scope of work. While a substantial amount of data was collected, the majority was generated by the University.

Several deliverables required the creation of new data. First, locating each mine required geo-referencing of the 6-month mining maps; as well as digitizing the extent of mining and locations of all features undermined. Second, creating the overburden maps required digitization of mining outlines, surface features and structure contours from the geo-referenced maps.

Overall, the University created data layers in seven categories: "Geo-referenced Maps," "Mining Outlines," "Surface Features," "Contours," "Overburden," "Buffers," and "Stream

Observations." The following sub-sections explain the methodology for the creation of this data as well as how existing data was integrated during the process.

II.E.1 – Geo-referenced Maps

This section explains the process utilized in geocoding the mine maps for this analysis. The first task was to determine the location of mines operating during the study period. As mentioned in Section II. C. over 900 6-month mining maps were acquired in hardcopy (paper) format from the PA DEP CDMO. A subset of these maps were scanned to create usable digital images. The method to geocode these scanned images into their location on the earth is called geo-referencing.

Folders obtained for each mine have sub-folders by year of operation during the assessment period that contain sheet maps and, if present, generally one copy of an index map. For example, 20 maps might be acquired for one mine that include an index and different versions of each sheet map over time. This results in a large number of maps with similar information. Since the University needed to identify the extent of mining and surface features on these maps deciding which version of the sheet maps to use posed a challenge.

The University developed a protocol for the most recent versions of all index and sheet maps to be entered into the geo-referencing stream. The analysis of those 900 maps led to the subsequent geo-referencing of 317 6-month mining maps for the assessment period. The average number of maps geo-referenced per mine for longwall mines was 16 maps while the average number per mine for room-and-pillar and room-and-pillar with pillar recovery mines was five maps. The protocol further stipulated the desired accuracy of the geo-referencing process. During the process the scanned images were loaded into the software application and referenced to base data such as roads and streams using control points. The protocol required a minimum of three control points to reference an image, two to pin the image to a plane, and a third to disable rotation. The application tool, called the Geo-referencing tool, calculated a root mean squared error (RMSE) for all control points in relation to each other. The protocol set the minimum number of control points at four and the largest allowable RMSE at 10-m. Figure II-1 shows the distribution of the RMSE and the mean number of control points used for each mine in the study area. The blue lines represent the mean number of points and red lines represent the average RMSE. As evidenced by the graph, mines with a higher average number of points tended to have a lower average RMSE. One outlier was Dooley Run which had nine control points and a RMSE of 9.6-m (meters). This is typically the result of divergent data points on the input maps. The average RMSE and average number of points for all 6-month mining maps were 3-m and 6.8, respectively. This meant the average horizontal accuracy for any 6-month mining map in the University GIS Database (UGISdb) was 3.0-m.

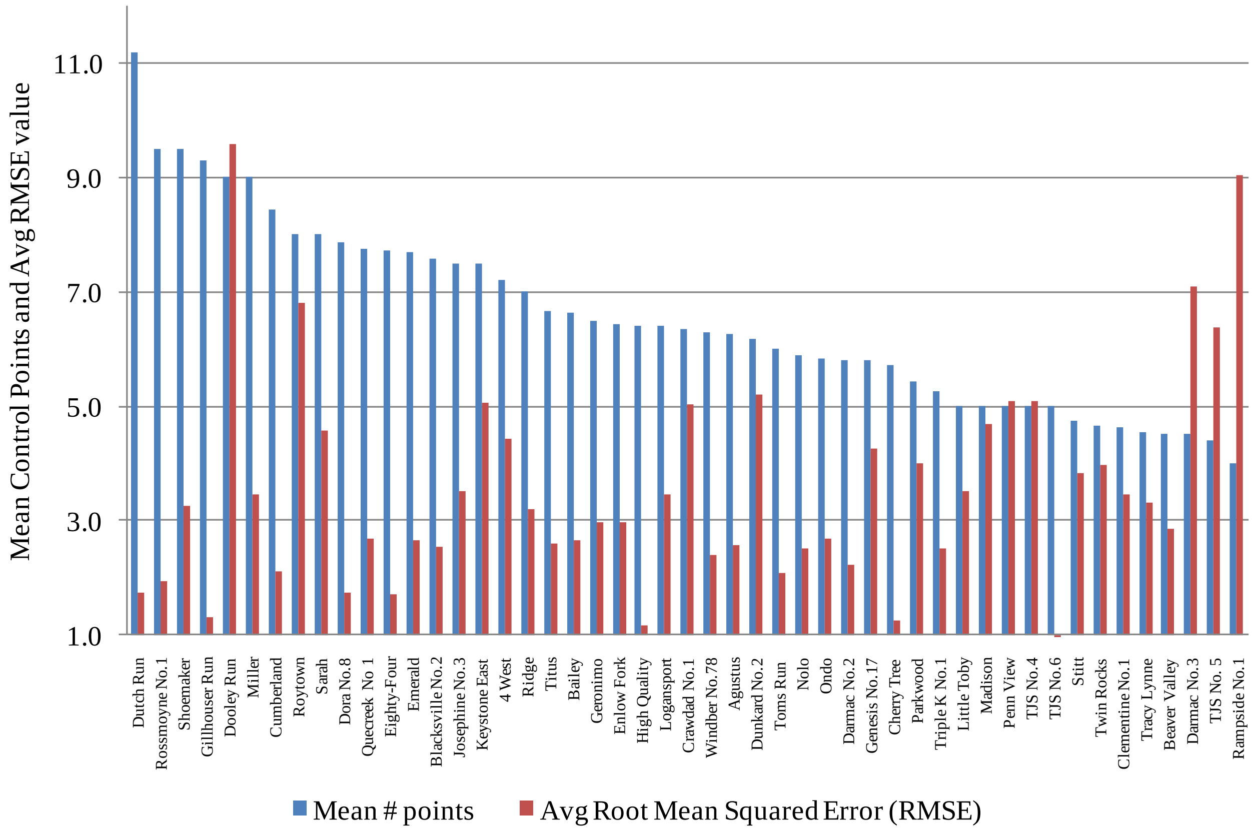

Figure II-1 - The mean number of control points and average root mean square error (in meters) for each mine in the study area. The total number of maps included here is 317.

II.E.2 – Mining Outlines

Once the location of all the mines was determined and a series of 6-month mining maps were geo-referenced, the University began the task of determining the extent or amount of mining that occurred during the assessment period. Refer to Section III for a listing of the acre per mine.

The extent of mining was generated by digitizing the outline of mining from the geo-referenced 6-month maps, and obtaining the outlines in AutoCAD or preferably, a shapefile format was received from a mine operator. All operator data was verified using the 6-month maps. In total, 24 mining extents were obtained from the respective operator, and the remaining 26 mines were digitized from 6-month mining maps.

II.E.3 – Surface Features

The most difficult task proved to be rectifying the BUMIS inventory with that of the GIS feature inventory. Features are defined within this study as properties, structures, and water supplies. In brief, this process included joining the BUMIS inventory to the spatial datasets that were generated from, or verified using, the 6-month mining maps.

A 200-ft buffer was created around all areas mined during the 3rd assessment period. The designation of 200-ft was a product of both the use of a 200-ft buffer by the CDMO for inventory purposes and the variance in other inventorying zones such as the 100-ft buffer for structures and the variable RPZ buffer for water supplies. The 200-ft buffer standardized the management of several feature types and other data layers to a uniform boundary limit.

The 200-ft buffer used in this process served as the perimeter of the GIS inventory with the exception of features with reported effects. A reported effect was any feature with a corresponding problem file indicating a possible mining related impact during the assessment period. The final totals for GIS features within 200-ft of mining or having a reported effect were: 3,587 properties, 3,735 structures, and 2,789 water supplies.

The remainder of this section details the process for digitizing and joining BUMIS feature spreadsheets with their respective spatial counterparts. There were many surface features on the 6-month mining maps that were within 200-ft of mining, yet not included in the inventory extracted from BUMIS; at the same time, there were many surface features in the BUMIS inventory that were outside 200-ft buffer. These discrepancies were rectified within UGISdb.

II. E.3.a - Properties

The University, using the property lists generated from BUMIS and the 6-month mining maps, identified properties within 200-ft of the extent of mining for each mine in the study. As mentioned above, properties listed in the BUMIS inventory that are located outside 200-ft from active mining were not included in the joining process unless the property listed a reported effect. All properties within or intersecting the 200-ft buffer were inventoried. The identification process found approximately 370 new properties not listed in the original BUMIS inventory. After a second round of queries, 175 of the new 370 properties were found in BUMIS and more detailed information was present in the UGISdb. The remaining 195 properties were added to the UGISdb, but lacked information commonly found in BUMIS such as property owner, feature use, etc. Reasons for the lack of information in BUMIS regarding the remaining 195 properties found within 200-ft of active mining remains unknown.

Once the two sources were evaluated, a further step was required to merge the two datasets into one. This process began with the feature shapefile and the BUMIS inventory spreadsheet for each mine. A unique identifier was added to each input and used as a common field to append the non-spatial tabular data into the spatial attribute table. Next, a record matching algorithm was applied to the spatial dataset and the non-spatial data is appended. This result was then exported to a new dataset and given a name to show it was successfully merged. This process appended the basic information derived from BUMIS, i.e. Owner, County, Feature Type, etc. Spacial data (maps) can now be queried by this non-spatial set of criteria.

In many cases, a second merge was required to append information related to reported effects for each feature, i.e. claim ID, occurrence date, resolution, etc. A unique identifier was placed in the new spatial dataset. Additionally, a second spreadsheet containing information regarding reported effects included this same identifier, the BUMIS Problem ID. Finally, a record matching algorithm was applied to the dataset, the new tabular information was appended, and an updated dataset was exported which had both the mine information and information regarding reported effects.

Limitations of the property dataset included the accuracy of property boundaries and inclusion of all properties present in the mined area. First, since the extent of input information for 50-pct of the mines was limited to the 6-month mining maps, the property boundaries only extended to the borders of these maps which in many cases do not represent the real boundary of the property. Second, if a property was subdivided toward the end of the assessment period, it could have been missed in the spatial inventory due to temporal conflicts between the dates of the BUMIS inventory, dates of the 6-month maps being used, etc.

II. E.3.b - Structures and Water Supplies

Once the property inventory was complete, the task of locating the remaining features, structures and water supplies, within each property boundary was the next step in the progression. The 200-ft buffer was also added, and any structure or water supply, without a reported effect, located outside the buffer was eliminated. A similar process, as described for the property identification, was used for these features. Lists generated from the BUMIS inventory were compared to lists generated from a spatial inventory using 6-month mining maps and operator provided data. Discrepancies found during this cross-referencing were resolved.

As time elapsed between the BUMIS inventory, in the fall of 2008, and the completion of the study, a discovery query tool was used to find all reported effects that occurred during the 3rd assessment that had not been entered into BUMIS by the end of 2008. These reported effects were referred to colloquially, by the University, as Blue Features.

In some cases, the blue features were located far from mining or out of the extent of the 6-month mining maps, and in others, the features were in the original spatial inventory without appended BUMIS information or were in the original BUMIS inventory but without a reported effect. A total of 365 blue features, or new reported effects, were found using the query tool. Of these 365 blue features 332 were assigned spatial locations and 33 were added as attributes in tabular form to the University GIS Database. The methods for locating the blue features with reported effects were as follows:

- Features that were found within the original inventory(s) were updated with new claim/problem information;
- Features not found in the original inventory were first located based on a property number, if present; or
- Using Topologically Integrated Geographic Encoding and Referencing system (TIGER) line files created by the U.S. Census Bureau, or
- Searching Google Maps web-based mapping application for the address listed in the dataset, and
- Once found, the information was updated in the UGISdb.

II.E.4 – Contours

Structure contours in this study included overburden, coal, and surface contour layers. These datasets were digitized from 6-month mining maps, obtained in AutoCAD, raster format from an operator, or taken from the PA portion of a regional coalbed assessment (Anon, 2009). There were a total of 14 overburden contour datasets, 36 coal contour datasets, and 2 surface contour datasets. Table II-2 shows the data sources for each category. Surface datasets were not included in the UGISdb. However, University generated overburden contour layers were included in the UGISdb. See sub-section II.E.5 for a description of the overburden creation process.

Table II-2 –Three categories of contour and their respective data sources.

Category	Digitized from Geo-referenced Map	Obtained from Operator	Used Chapter D Coal Contour (USGS)
Overburden	1	13	0
Coal	24	11	1
Surface	1	1	0

II.E.5 – Overburden

The creation of an overburden map for each mine was an important task. The resulting overburden dataset was used as an input for the generation of the RPZ buffer and was referenced as a key source of all analysis in the report. Two main processes were used to generate the resulting overburden maps. Each process is defined in the following paragraphs.

The processes were divided by the contour input. Since the database included both coal contours and overburden contours, there were parallel processes used to reach the desired result, a final overburden contour dataset accurate to 500-ft beyond the edge of mining.

The process used for an overburden contour input was interpolated using spline interpolation method. This method, developed by Hutchinson (1989), is still the only contour-based interpolation method in use. The tool, found in the Spatial Analyst module of ArcEditor, was used to interpolate the overburden contours in line form to a continuous raster surface with 10-m pixels (cells). Each contour was measured using units of feet resulting in the surface cells storing measurements in feet. A buffer of 500-ft was used in the interpolation process; therefore, the accuracy of the interpolation is bound by this distance.

In a parallel process, if coal contour dataset was used as input, there was an additional step to create the overburden map. The coal contour was interpolated to a continuous surface as described above. Meanwhile, the two National Elevation Dataset DEMs of the surface were re-sampled from 30 to 10-m pixels. The final step was to subtract the interpolated coal elevation raster surface from the modified DEM raster surface with the difference image representing the overburden surface. Once a raster surface was derived for each mine, each surface was used to generate new overburden contours extending 500-ft beyond mining.

One significant limitation for this dataset was the provenance of coal contours in both the 6-month mining maps and AutoCAD files. Knowledge of the National Elevation Dataset origin could be combined with knowledge of contour datasets to develop an estimation of temporal resolution. Due to the absence of this key piece of information an accurate temporal resolution for the generated overburden datasets could not be identified.

An estimation of horizontal accuracy for the resulting overburden contour datasets were made for coal contours digitized by the University (approximately 24 mines). An estimation of this error can be found by adding the pixel size of University generated overburden surfaces (10-m) and the average geo-referencing error (3-m) resulting in a maximum of a 13-m horizontal error. A more detailed estimate of error can be calculated for each mine using the average geo-referencing error for each specific mine. Other datasets were estimated to have at least 10-m error, yet the real error could be larger. One mine had no available contour information so data from Chapters C and D of Resource Assessment of Selected Coal Beds and Zones in the Northern and Central Appalachian Basin Coal Regions (Anon, 2000) was used. These data were generated using 100-m pixels and the horizontal error was estimated as high as ¼ mile (approximately 400-m). A statement of vertical accuracy cannot be made due to the lack of provenance for all contour data with the exception of Chapter C and D data which estimate a 30-m vertical accuracy.

II.E.6 – Buffers

Several buffers were used in the study to determine relationships based on proximity, create a boundary for analysis, or model an existing regulatory boundary. These buffers included both uniform and variable distances, and extended both outside or around a feature and inside or within a feature (Figure II-2).

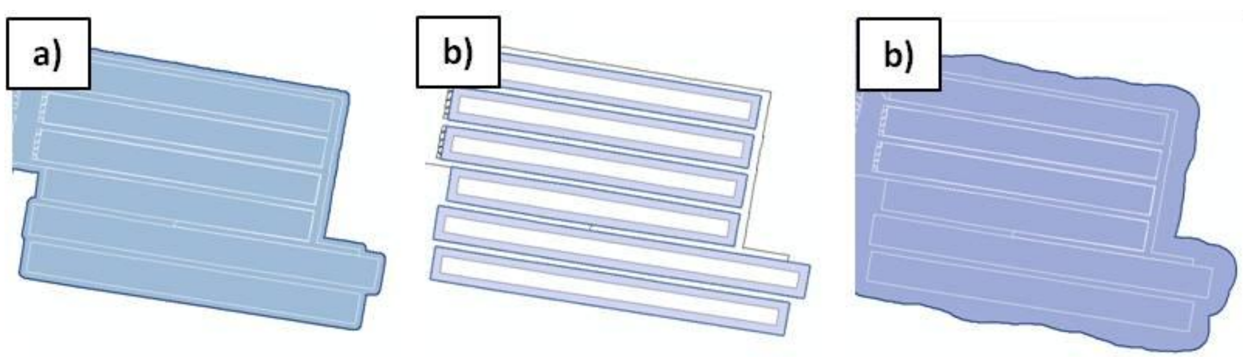

Figure II-2 – The uniform, variable, inner and outer buffers used in the study; a) Uniform outer 200-ft buffer, b) Variable inner – Quarter width, and c) Variable outer – RPZ.

II. E.6.a - Uniform Distance Buffer

Two uniform distance outer buffers were used for all mines in the study. A 200-ft (Figure II-10a) and 500-ft buffer were generated from the mining outline and used as a boundary for feature inventory and analysis, and overburden generation, respectively.

II. E.6.b - Quarter-Width Panel Buffer

A variable, inner buffer was generated for each longwall panel based on the following equation (Panel Width x 0.25). The resulting buffer was generated inside the polygon panel feature (Figure II-10b). This buffer was used to determine each water supply and structure's position along the longwall panel and inference about proximal relationships between the surface feature and its position over mining.

II. E.6.c - RPZ Buffer

A variable, outer buffer was generated based on the depth of mining (Figure II-10c). Two equations were used to create the buffer: (Overburden value x Tan (35-deg x (π/180)) and (Calculated RPZ Distance / 3.281). For a full discussion of the purpose and use of the RPZ boundary by the PA DEP, refer to Section VI.C.

The creation of this variable, outer buffer was an effort to model the complex boundary used by the PA DEP. The input dataset was the overburden map discussed in Section II.E.5. The raster surface was then converted to a polygon shapefile where each polygon stored an overburden value. This new polygon shapefile was clipped to the edge of mining.

The equations listed above were applied to the overburden values at the edge of mining and stored as an attribute. The variable buffer was created based on this calculated attribute. The resulting dataset included many small buffers around each polygon and a dissolve procedure was used to generate visual clarity. Note that equation two was necessary because the spatial reference used by the dataset was measured in meters and the input overburden dataset is measured in feet; therefore, the calculated RPZ distance used in the buffer generation must match the units of spatial reference.

II.E.7 – Stream Observations

The University was tasked with surveying streams with reported effects or for control conditions. The goal was to obtain data for tabulating a Total Biological Score for impacted streams. For more information about stream assessments and calculated Total Biological Scores see Section VII.

II. E.7.a - GPS Data Collection

During each surveying trip the University Team used a Garmin ETrex GPS device to accurately capture each sampling location. The status of sections of each stream walked was caterogrized as flowing, dry, or intermittent. This data was imported into Garmin MapSource software and exported to an AutoCAD Drawing Exchange Format File which was converted to an ESRI shapefile for use in the University GIS database. The GPS data points were collected in the World Geodetic System of 1984 geographic coordinate system which is the default coordinate system used in most GPS data collection. After extraction from the device, the GPS data points were transformed to the North American Datum of 1983 and projected to the UTM Zone 17 North projection in accordance with study standards.

II. E.7.b - Stream Segments

After the GPS data was extracted and placed in the UGISdb, the University digitized segments between the captured points following the streams on an existing stream dataset called Net Streams 1998. The resulting dataset included all sections of surveyed stream and their status on the date of observation. Maps showing these observations are included in Appendix D.

II.E.8 – Land Use

The digitized stream segments discussed in the preceding section were used to create a layer of data showing the associated land use within 100-ft of the stream. Each surveyed segment was given a 100-ft buffer and this buffered area was used to select only overlapping land use information for that area. For a full discussion of the results of this analysis refer to Section VII.

II.E.9 – Wetland Maps

Wetland maps for several mines were obtained from one longwall operator. Of these, 21 were geo-referenced with an average of 3.27 control points and 1.69-m RMSE. A more accurate estimate of the area of wetlands in these mines were not possible as several maps marked a wetland as a point, rather than defining the encompassed area; and maps with more accurate delineations included few features for geo-referencing. A second limitation was the age of the only national wetland dataset. The National Wetland Inventory, created in 1992, was no longer representative of the number of wetlands currently present in the study area. Refer to Section IX for more discussion on wetlands.

II.F – University GIS Database Structure

The UGISdb was structured as a functional collection of discrete datasets organized by mine. The contents included AutoCAD data, the geo-referenced 6-month mining maps, University generated spatial data, University generated BUMIS spreadsheets, and the geo-database of existing data clipped to the extent of the county where the mine operated. The benefit of maintaining the dataset in this structure is its portability. With this strategy, each mine's dataset is manageable, but if it had been organized by feature type, any assessment of a particular mine would have required copying data from many different folders and sharing data that would have been time consuming. The following section defines the file structure of the database, and associated attributes assigned to specific layers.

II.F.1 – File Structure

In the mine information portion of the database, a folder for each mining type (i.e. Longwall, Room-and-pillar, and Room-and-pillar with Recovery) assessed in the study provides the highest level of the database. Each mining type folder includes folders for each mine classified as that type in this study period. Each mine folder included all data collected or created by the University.

Other categories of data in the database include Streams and Wetlands. These data extend beyond the boundaries of a single mine so the data is managed at the feature level. Folders for these features are the highest level of the database along with the mine information data. Each folder is then organized by data type similar to the mine information dataset. Figure II-3 compares the size of all datasets, mine information, streams, and wetlands.

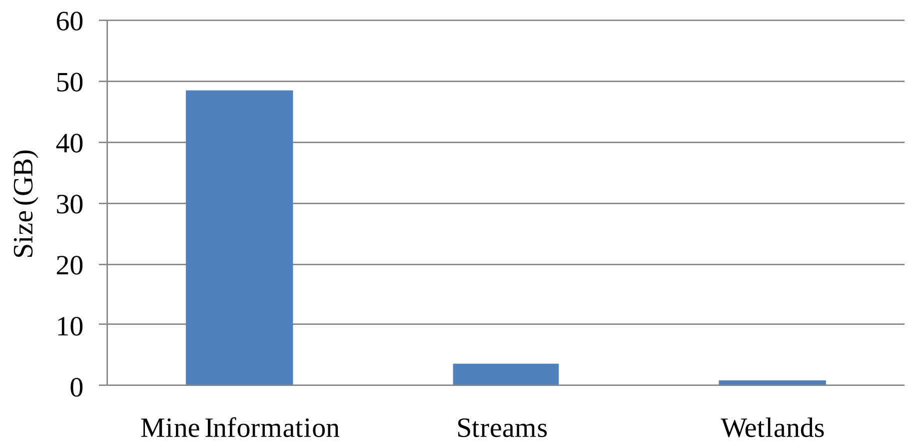

Figure II-3 – Comparison of the size of the three highest level datasets, Mine Information, Streams, and Wetlands. Mine information is the largest dataset with 48.5-GB, the Streams dataset contains 3.5-GB, and the Wetland dataset is much smaller at 0.8-GB.

II.F.2 – Attributes

Every dataset has a set of attributes. Attributes are non-spatial information that describe a spatial feature. There are approximately 22 feature files in each mine dataset. Each dataset includes specific attributes for analysis. A general list of primary attributes for Features, Mining Outlines, Contours, Overburdens, and Buffers follows.

II. F.2.a - Features

The surface features are a core product of the study. These features are used to assess the overall impact of underground bituminous coal mining within the study area. General attributes for the three generated feature layers, properties, structures and water supplies are as follows: Property Number, Property System ID, Unique Identifier, Property Owner, County, Feature ID, Feature Number, Feature Type, Feature Use, Problem ID, Claim Number, Cause, Occur Date, Interim Resolution Date, Final Resolution Date, Resolution Status, Mine Name, Operator, Topographic Location, Overburden, Surface Elevation, Panel Location, Coal bed, and Area in Square Feet.

II. F.2.b. - Mining Outlines

The mining outlines are used to calculate the acres of mined coal, serve as the boundary to base the 200-ft buffer inventory extent, and provide positional information for reported effects to water supplies and structures. General attributes for these datasets include: Mine Name, Operator Name, Coal bed, Status, Mining Type, Acres, and Area in Square Feet. Longwall Panel datasets included extra attributes such as: Panel Number, Panel Length, Panel Width, Year Complete, Start Date and End Date.

II. F.2.c - Contours

The contour layers are used to generate elevation surfaces for both coal and overburden elevations. General attributes in these datasets include: Mine Name, Operator Name, Coal bed and Elevation Value.

II. F.2.d - Overburden

Overburden datasets are used for analysis of water supplies and structures and provide an analysis of the average depth of mining for mines operating during the assessment period. The raster overburden datasets do not have multiple attributes. The raster data structure stores a single attribute, in this case an elevation value in each cell composing the grid. The raster overburden datasets were converted to shapefile format and more attributes were added to this dataset. General attributes for the overburden shapefile include: elevation value, and calculated RPZ distance.

II. F.2.e - Buffers

Several buffers were employed in the study. A uniform buffer of 200-ft was used to inventory all surface features, a quarter-width buffer to determine panel position for water supplies and structures, and a custom buffer for the RPZ distance from mining. General attributes for the 200-ft buffer include: Mine Name, Operator Name, and Buffer Size. General attributes for the quarter-width buffer include: Mine Name, Operator Name, Panel Number, and Buffer Size. General attributes for the RPZ buffer include: Mine Name, Operator Name, Buffer Distance, and Dissolve.

II.G – Metadata Creation

Due to the immense effort put forth to create these datasets, the University places a high priority on enabling the efficient use of the data in the future. Metadata was added to all datasets to ensure the proper information about purpose, geo-processing, and other data manipulation is transferred to future users. A formal definition lists the main components of metadata as properties and descriptions. The properties are pieces of information populated automatically by the GIS software such as the spatial reference and geo-processing history while descriptions are human entered information explaining access, use, purpose, current location, and publication date of a dataset. For any data to be truly useful, it is important to have information about the data including what the data is used for, how it can be used and accessed, and especially, how it was created.

The metadata standard chosen for this study is the Content Standard for Geographic Metadata developed by the Federal Geographic Data Committee (FGDC). This standard was developed for and is used predominately in the U.S., but several data providers in Canada have also employed it for metadata design.

This metadata standard is supported by the ArcCatalog module of ArcEditor. To access the metadata for these datasets, select the FGDC Classic as the metadata viewing format within the module and then select the dataset for which the information is needed. The FGDC metadata standard includes the following categories:

- Citation Information,
- Publication Date,
- Title,
- Abstract,
- Purpose,
- Time Period of Content,
- Spatial Coordinates,
- Keywords,
- User Constraints,
- Attribute Information, and
- Contact Information.

II.H – Summary Observations

Data collection proved to be an on-going process. The amount of data collected resulted in databases of 900-MB and 53-GB. Existing data was obtained from many providers including local, state, and federal government entities, non-profit organizations, and mine operators. Efforts to gather existing data, geo-referencing and feature rectification proved the most time consuming tasks. There were several new data products created during the 3rd assessment period, such as the generation of overburden maps, RPZ maps, and quarter-width panel buffers.

All data in the UGISdb is transformed into the reference standard for the study. Based on its accuracy for the entire region under study the reference standard is viewed as the best choice; however, as models of the earth and new projections see continued development, this could most certainly change in the future.

The utility of the database structure proved itself time and again as a flexible and portable collection which enabled many to utilize each dataset and understand its simple structure. While the database structure proved efficient, the task of data management included many hurdles.

The creation of metadata in the UGISdb ensures the data created during this assessment has the potential to be used for many years to come. The methodology and database presented in this report will both inform researchers of potential obstacles, as well as aid them in amassing large datasets in future assessment periods.