

# **DEVELOPMENT OF A CARBON MANAGEMENT GEOGRAPHIC INFORMATION SYSTEM (GIS) FOR THE UNITED STATES**

## **Technical Progress Report**

Reporting Period: October 1, 2002 – September 30, 2003

Principal Author:  
Howard J. Herzog

Issue date: March 2004

DOE Award No. DE-FC26-02NT41622

Massachusetts Institute of Technology  
Energy Laboratory  
77 Massachusetts Avenue  
Room E40-447  
Cambridge, MA 02139-4307  
Telephone: (617) 253-0688  
Fax: (617) 253-8013

## **Disclaimer**

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

## Abstract

The Lab for Energy and Environment (LFEE) at MIT is developing a Geographic Information System (GIS) for carbon management. The GIS will store, integrate, and manipulate information relating to the components of carbon management systems. Additionally, the GIS can be used to interpret and analyze the effect of developing these systems. In the first year of this three year project, we focused on two tasks: (1) *specifying the system design* - defining in detail the GIS data requirements, the types of analyses that can be conducted, and the forms of output we will produce, as well as designing the computer architecture of the GIS and (2) *creating the “core” datasets* - identifying data sources and converting them into a form accessible by the GIS.

## Table of Contents

<b>Disclaimer</b> .....	2
<b>Abstract</b> .....	3
<b>Executive Summary</b> .....	5
<b>Experimental</b> .....	6
<b>Results and Discussion</b> .....	7
GIS Design and Implementation.....	7
CO <sub>2</sub> Sources .....	11
CO <sub>2</sub> Sinks.....	15
<b>Conclusion</b> .....	25
<b>References</b> .....	25

## Executive Summary

During this past year, our work has focused on Task 1 and 2 of our proposal. Summaries follow below:

**Task 1, System Design:** In this task, we are defining the GIS data requirements, the types of analyses that can be conducted, and the forms of output we will support. In addition, we are designing the computer architecture of the GIS.

The GIS computer architecture includes the decisions on computer hardware and software. Our basic architecture is based on off-the-shelf computers and software, giving us more time to work with the data and added-value tools. We are developing our GIS with multiple computers. Programs that we have decided to use can communicate with each other over the network, working together to respond to the GIS user. Starting with a network of multiple computers also allows us to support future growth of the GIS's functionality and increased user load because new computers can be added to the system without major changes to the current design. Key milestones include:

- Meeting with MIDCARB to learn from their current practices
- Two personal computers purchased as the initial network of computers
- Database development computer
  - Oracle Database
  - ArcSDE for Oracle
- User interface/front end computer
  - ArcGIS

The data requirements define the required characteristics of the various parts of a carbon management system. The data requirements are closely linked to the analysis that we expect to conduct. These include:

- Data requirements for primary carbon dioxide system: sources, transportation infrastructure and sinks.
- Data requirements for factors that may modify costs in the system: geography, topography.
- Analyses
  - Identify distribution of sources and sinks and areas of high potential for carbon sequestration
  - Estimate capital cost and operating cost for capture, transportation, and injection at individual locations
  - Identify viable source/sink pairs
  - Create carbon sequestration cost curves

We have produced a working prototype that incorporates the first two bullet points above. We are currently working on items in the third bullet point.

**Task 2, Core Datasets:** This task consists of the identification and incorporation of data into the GIS. While this will be a continuing process, an initial set of data has been gathered into the GIS so we can begin basic analyses. These include:

- CO<sub>2</sub> sources

- Power Plants: EPA's eGRID
  - <http://www.epa.gov/cleanenergy/egrid.htm>
  - Data download from website
- Industrial facilities: Ecofys dataset on worldwide sources.
  - *Building the cost curves for CO<sub>2</sub> Storage, Part 1: Sources of CO<sub>2</sub>*, July 2002
  - Data download from CD
- CO<sub>2</sub> sinks
  - Brine Aquifers: University of Texas, BEG's Brine formation database
    - <http://www.beg.utexas.edu/enviroq/ty/co2seq/dispslsalnt01.htm>
    - Data download from website
  - Oil and Gas Reservoirs: DOE's Gas Information System (GASIS)
    - <http://www.netl.doe.gov/scng/projects/model/r-d/rdp28139.html>
    - Data download from CD
  - Coal Seams: USGS's Coal Fields of the United States
    - <http://www.nationalatlas.gov/coalfdm.html>
    - Data download from website
  - Ocean depths: NGDC five-minute topography (ETOPO5)
    - <http://www.ngdc.noaa.gov/mgg/global/setopo.html>
    - Data download from website
- Supporting data
  - State and county boundaries: USGS
  - County population: Census
  - Urban and metropolitan areas: USGS

A paper on this topic was presented at the Second Annual Carbon Sequestration Conference – “Analysis and Status of Carbon Management GIS Data,” by D. S. Cheng, A. M. Smith, T. E. Curry, H. J. Herzog

### ***Supporting Activities:***

- Visited the MIDCARB project in Lawrence, KS (October 2002) to coordinate our effort with what is being done at MIDCARB.
- Attended the Geological Society of America (GSA) Conference in Denver, CO (October 2002) to investigate how GIS is being used in conjunction with geology.
- Attended the Coal-Seq II Forum in Washington, DC (March 2003) to better understand how coal can be characterized and used for sequestration. This will allow us to better design the coal features in the GIS.
- Attended the Second Annual Carbon Sequestration Conference in Alexandria, VA (May 2003). A paper on this project was presented.

## **Experimental**

This project involves computer modeling and there is no laboratory work associated with this project.

## Results and Discussion

The Lab for Energy and Environment (LFEE) at MIT is developing a Geographic Information System (GIS) for carbon management. The GIS will store, integrate, and manipulate information relating to the components of carbon management systems. Additionally, the GIS can be used to interpret and analyze the effect of developing these systems. The first section below summarizes the design of our Carbon Management GIS.

In order to model carbon management options, the GIS must contain information on CO<sub>2</sub> sources and CO<sub>2</sub> sinks. A variety of public datasets have been examined in search for suitable information. The second and third sections below describe the datasets that have been evaluated and their utility to the MIT GIS. Data utility has been divided into three different factors: coverage, characterization, and timeliness. Coverage expresses how many of the locations in the area of interest are contained in the dataset. Characterization represents the range of information available for each location. Timeliness measures the relevance of the data to the time-frame under consideration.

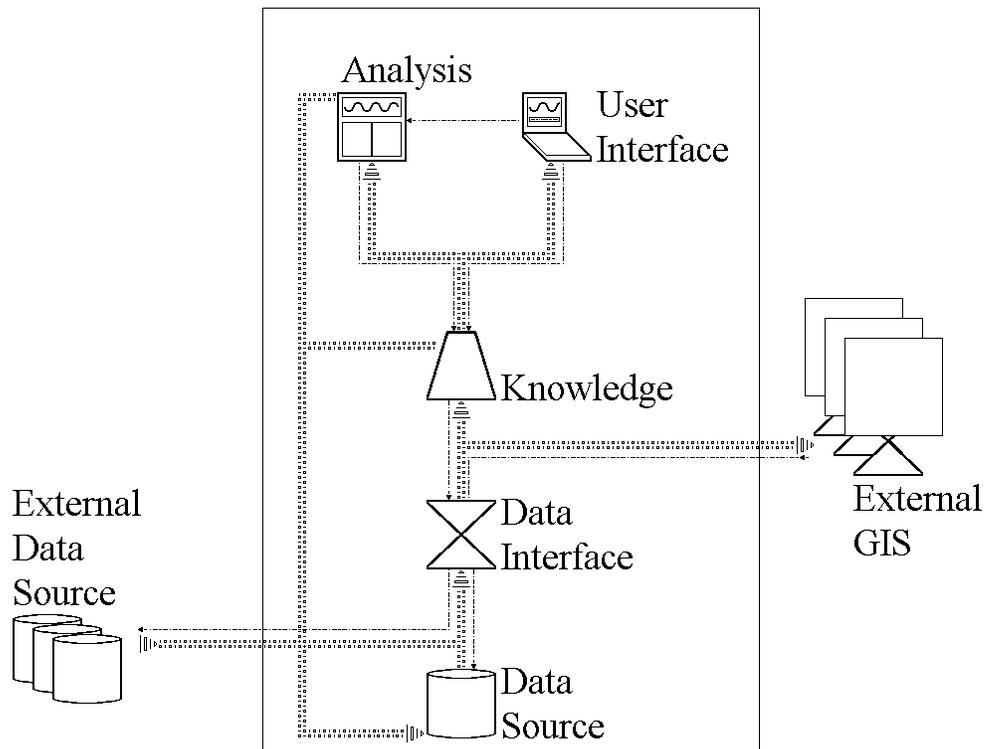
### *GIS Design and Implementation*

#### *System Overview*

Figure 1 is a diagram of the layers in the GIS as well as the control and data flows. Thin dotted lines represent control signals and double lines represent data flows. Layers of the system are labeled, representing distinct modules. The local system is enclosed in the solid box.

The layers of the system are as follows:

- The User Interface Layer (UIL) provides access to the system in a human usable format. Based on this interaction, the UIL sends control signals to the Analysis Layer (AL) to initiate analyses and to the Knowledge Layer (KL) to request information.
- The AL runs analyses and models. The data required for the analyses are retrieved from the KL, and results are stored into the Data Source Layer (DSL).
- The KL integrates data into coherent sets of information based on user requirements and available data. The KL receives requests from the UIL or AL, then collects data from the Data Interface Layer (DIL) and integrates the data that is relevant to the request. These results can be stored in the DSL for future use.
- The DSL represents all the databases and sources of data available to the GIS. It provides data based on queries from the DIL.



**Figure 1: System Schematic**

Although each module may be developed and hosted at a different site, the boundary of the local system is measured by use of a common DIL and KL. External data sources are accessed by the DIL but not managed by the local system. External GIS do not use the DIL and KL for context mediation and integration, but may request data from the DIL.

The GIS at MIT is implemented on two consumer grade desktop PCs. Each of these machines is running primarily standard software.

Oracle and ESRI products - Oracle database, Oracle application server, ESRI ArcGIS, and ESRI ArcIMS - have been chosen to support the reproducibility of the system. These software products are heavily used in the CCS community and at MIT. Therefore, other groups will be able to reproduce techniques developed at MIT in their own systems. The software code developed to incorporate information integration technologies are written in the Oracle database in the PL/SQL and Java programming languages. Some analysis programs have also been written for ArcGIS in Visual Basic, the scripting language used by that software package.

## ***User Interface Layer***

- *Goal:* Provide an interactive environment that can accommodate all of the potential users that include technical analysts, policy decision-makers, and the public.
- *Goal:* Provide a non-intrusive interface by having minimal software and processor requirements.
- *Design:* Handle user input to trigger queries to the KL and commands to the AL.
- *Design:* Display map and information screens that are graphical and straightforward.

In order to fulfill the goals of an accommodating and non-intrusive user interface, we have implemented an internet website. The website is developed in the Apache web server within Oracle 9i Application Server. These software packages form the basic web server, onto which the ESRI ArcIMS software is added to provide graphical interfaces and map displays. These off-the-shelf products allow us to quickly prototype a User Interface. Additionally, the requirements on the user side are minimal because all of the computation is done on the server side. A user of the GIS only needs a web browser and internet connection.

ArcIMS display maps and data that are stored in a database or on the server's file system. Specific scripts (ArcSDE) are required by ArcIMS and have been added to CCSPs database to handle control and data connections between the KL and ArcIMS.

The interface to send controls to the AL has not yet been implemented. These will be handled by programs in the web server that will take user commands and run the appropriate analysis program.

## ***Analysis Layer***

- *Goal:* Enable tools and models to analyze major components of a CCS system: capture feasibility, transport routing, sink selection, and cost estimates for each of the components.
- *Goal:* Allow extensibility of analysis modules in support of system design goal.
- *Design:* Capture feasibility requires measurement or estimation of emission quantity and concentration, and requirements of capture technologies.
- *Design:* Transport calculations require the location of a source and sink pair, and factors that modify costs such as terrain, right of ways, and transport options.
- *Design:* Sink selection requires information on the reservoir characteristics: depth, thickness, permeability, porosity, pressure, and temperature.
- *Design:* Analysis tools should communicate with the database but be implemented in any programming tool appropriate.

Analyses are developed using a variety of programming languages and tools. The GIS can incorporate these tools into the system as long as the results are stored into the database for display and further analysis.

For example, the sink injection model runs in ArcGIS. The model takes the reservoir characteristics as input in the form of two-dimensional grids, stored in ArcGIS raster files. It then runs the injection costing algorithm [Heddle, 2003] that was developed at MIT on the grids

to calculate the estimated cost of drilling wells and a per-ton cost for injection. Other analysis programs have been written for the Oracle database, including calculations of emissions from sectors in each of the regions of the U.S.

### ***Knowledge Layer***

- *Goal:* Integrate data from different sources into collections of information that relate to the same real-world entity.
- *Goal:* Provide a single information interface for users and analyses, and supply integrated information to users in an understandable way
- *Design:* Integration is performed by building knowledge objects that represent the available information on the entity.
- *Design:* The knowledge objects will retain information on the data interfaces used in integration to enable tracing of data flows and data quality.
- *Design:* Correlate naming, labeling, and primary key conventions from the different data sources to locate related information.

The Knowledge Layer (KL) utilizes multiple Data Interfaces to gather data from multiple sources and integrates these data into coherent sets of information. A KL object represents the best information that is available for a certain topic.

For example, a KL object for power plants is designed by selecting the relevant data interfaces: eGRID, MIDCARB, and GESTCO. The data is correlated by determining a set of data fields that uniquely define a power plant in each of the sources and each field that represents the same power plant attribute. Each data source is given a quality rating and data for each attribute is retrieved from the source with the highest quality rating for use in the KL object.

### ***Data Interface Layer***

- *Goal:* Mediate context differences between sources and the local system through data conversion and translation.
- *Design:* Each type of data source will have an associated DIL module tasked with interpreting the data.
- *Design:* Translations between source and local context are centralized to allow reuse and avoid errors.

The DIL centralizes the services that are required to handle data source connectivity and context management. In order to connect to the variety of data sources that are needed by the system, several different methods of retrieving data are implemented. Context differences of source data are translated as it is brought into the system into the local system of units so that they can be

Data source connectivity is implemented through the use of database scripts and internet connections. Scripts are used to when the data source is provided by the data providers as computer files. The scripts extract the data from these files by converting the various file formats into a simple format and then loading it into the database. Internet connections are used when data is provided directly from remote databases. The Oracle database being used supports standard database transfer protocols used by data providers.

Context management is implemented by filtering the source data through database views. Meta-data is stored to keep track of the measurement units used by data providers for each data source. Additionally, the functions used to convert between different units are stored to enable reuse with different data sources. The database view is then created with the appropriate conversion function for each element of the data source.

### ***Data Source Layer***

#### *Data Source Goals*

- *Goal:* Build collaborations with data collectors.
- *Goal:* Understand current state of data sources available to research in CCS.
- *Goal:* Coordinate with data collectors to provide source data with improved information quality.
- *Design:* Support the goal of reproducibility with efficient, simple and interoperable database.

The Data Source is used to maintain process information needed by layers and to store the results of integration and analyses. It is implemented in an Oracle 9i database with additional ArcSDE scripts that are used by the ArcGIS analysis tool and ArcIMS web mapping software to interface with the database. The next two sections review some of the databases incorporated into the system.

### ***CO<sub>2</sub> Sources***

Point sources including power plants and industrial sources offer the most viable option for carbon capture because of the cost benefits from large scale capture. Power plants generate large volumes of CO<sub>2</sub> while many industrial facilities generate high purity streams of CO<sub>2</sub>. It is important to be able to characterize sources by their location, emissions, and factors that affect the costs of capture such as plant type and production method.

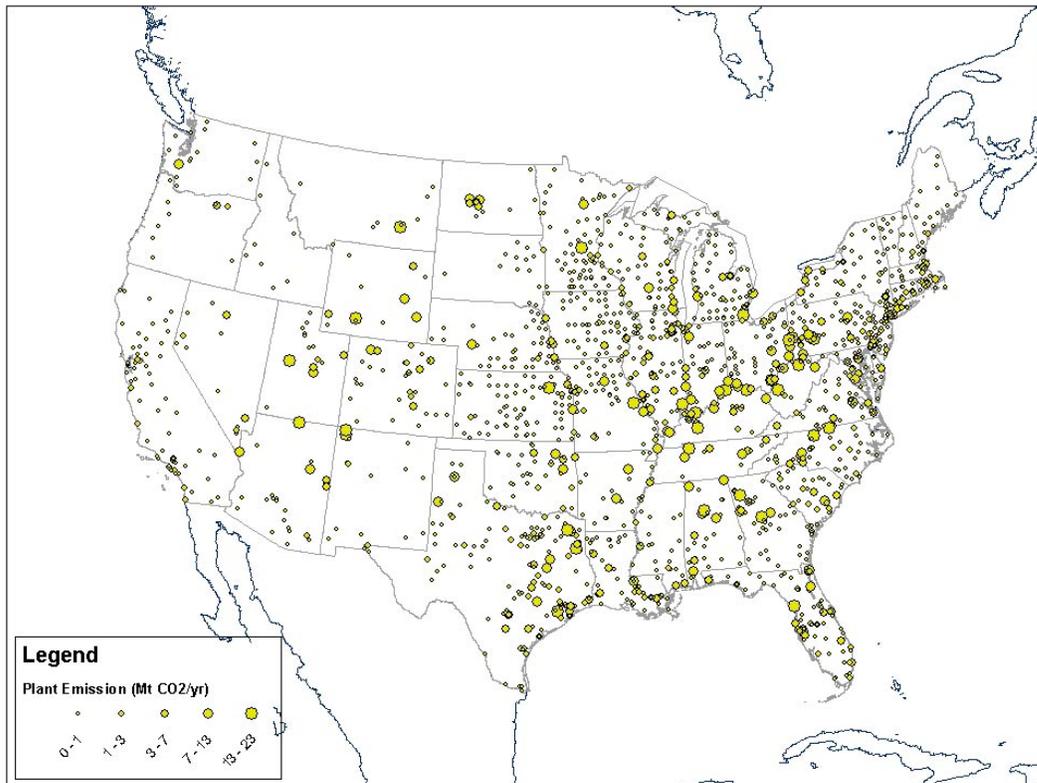
### ***Emissions and Generation Resource Integrated Database(eGRID)***

The Environmental Protection Agency (EPA) initially released the Emissions and Generation Resource Integrated Database (eGRID) in 2000, and recently updated the database in 2002. The EPA developed eGRID to support policymakers in decisions involving emissions and generation from the electric power industry.

eGRID integrates 24 data sources from the EPA, Energy Information Administration (EIA), and Federal Energy Regulatory Commission (FERC), some of which were confidential previous to the creation of eGRID. The database includes data on emissions (NO<sub>x</sub>, SO<sub>2</sub>, CO<sub>2</sub>), generation (capacity, resource mix), ownership, and location for over 4,500 power plants. eGRID also provides the data in several aggregation levels: boiler, generator, plant, company, state, and power grid. The data is distributed both as a stand-alone data viewer and as a set of Excel spreadsheets. The viewer can be used to run simple searches and queries on the data whereas the spreadsheets are available so that developers of more complex tools can access the data directly.

MIT is using data from the power plant aggregation level, and importing from the Excel spreadsheet to the GIS. The information on CO<sub>2</sub> emissions, resource mix, and location will be most important to the characterization of power plants as a source.

**Coverage:** eGRID contains data for nearly all electric generation units in the US and provides the geographic coordinates of the plants. The data covers plants throughout the US as shown in Figure 2 (Hawaii and Alaska not shown).



**Figure 2: eGRID Plant locations. Markers scaled by quantity of CO<sub>2</sub> emitted in 1998.**

**Characterization:** eGRID characterizes the plants by emissions, generation, ownership, and location. MIT has incorporated the power plant data for 1998. The database contains 142 fields describing the characteristics of each power plant. The types of data in the fields are as follows.

- Identification
  - Plant name, plant code
  - Plant type
  - Plant operator
- Location
  - State, county
  - Geographic coordinates
- Emissions
  - NO<sub>x</sub>, SO<sub>2</sub>, CO<sub>2</sub>, and mercury measured

- Measurements for total annual output, annual output rate per power output, and annual output rate per heat input
- Generation
  - Number of boilers and generators
  - Primary fuel
  - Generation and percent generation from coal, oil, gas, nuclear, hydro, biomass, wind, solar, geothermal, other fossil, and solid waste
  - Generation boiler capacity and name plate capacity
  - Heat input
- Ownership
  - Owner name, code, percent and type for the top 14 owners

**Timeliness:** The first version of eGRID, formerly known as E-GRID2000, was released in 2000. This version provided data for 1996 to 1998. In 2002, eGRID was updated with data for 1999 and 2000. Having current data on power plant emissions is important for developing carbon management systems because emissions levels change and new plants are introduced into the grid.

**Modifications:** The EPA spreadsheet is converted to a comma separated value (CSV) file. This CSV file is imported into the Oracle database. Additionally, the longitude field is stored as positive degrees west in eGRID. Since a GIS uses geographic degrees (degrees west are negative numbers), the longitude field needed to be negated.

### ***Clean Air Mapping and Analysis Program***

Another database considered was the EPA's Clean Air Mapping and Analysis Program (C-MAP). C-MAP contains emissions characteristics of the Title IV electric utility plants in the US. The database includes information on capacity, annual emissions of NO<sub>x</sub> and SO<sub>2</sub>, and monthly emissions of CO<sub>2</sub>.

This data is not being used in the GIS because the data found in eGRID represents a superset of the plants in C-MAP and has improved characterization.

### ***IEA PH4/9 Sources of CO<sub>2</sub>***

Ecofys prepared a report on worldwide CO<sub>2</sub> sources for the International Energy Agency Greenhouse Gas R&D Programme (IEA GHG). Ecofys provided MIT with a CD including the report entitled "PH4/9 Building cost curve for CO<sub>2</sub> Storage: Sources of CO<sub>2</sub>," and the source data supporting the PH4/9 report (PH4/9). This study focuses "on the location and size of large anthropogenic CO<sub>2</sub> sources."<sup>1</sup>

PH4/9 contains data for different types of CO<sub>2</sub> sources throughout the world. Data for the following types of production facilities were collected: ammonia, hydrogen, ethylene, ethylene oxide, oil and gas, refineries, iron and steel, power plants, and cement. The report describes the source of information and means of used for CO<sub>2</sub> estimation for each of the types of facilities.

---

<sup>1</sup> Hendriks, Chris, et al. *Building the Cost Curve for CO<sub>2</sub> Storage: Sources of CO<sub>2</sub>*. Ecofys. March 2002.

**Coverage:** Ecofys attempted to be as complete as possible in covering major sources of CO<sub>2</sub> worldwide. Data was collected from a variety of journals and reports totaling about fifteen thousand facilities worldwide for various industrial processes. For much of the data, facility locations were provided in terms of city or county and country. Ecofys attempted to find reasonable coordinates for facilities in order to use the data in a GIS.

Ecofys used the E-GRID2000 database to populate its power plant section. Zero emission renewable plants and smaller plants were removed from the database. eGRID also supplies coordinates for each power plant as noted in the eGRID section above. For the remainder of the facilities in the US, Ecofys queried the USGS Geographic Names Information System (GNIS)<sup>2</sup> with the city name associated with the facility. The documentation indicates that 96% of US facilities were assigned coordinates.

Table 1 shows an outline of the coverage information in the PH4/9 database. For each sector, the number of US facilities, the percent of these facilities with estimated coordinates, and the source of initial facility information.

**Table 1: Coverage Information for PH4/9 Database**

<b>Sector</b>	<b>US Facility Count</b>	<b>% with coordinates</b>	<b>Information Source</b>
Ammonia production	38	71	International Fertilizer Development Center (IFDC) 2000
Hydrogen production	87	90	Oil & Gas Journal (OGJ), Worldwide Refining Survey 1999
Ethylene	39	85	OGJ, Ethylene Report 2001 OGJ, Worldwide Construction 2001 ChemExpo, website
Ethylene oxide	13	85	ChemWeek, Website OGJ, Worldwide Construction 2001
Oil & Gas Processing	584	89	OGJ, Worldwide gas processing 2001
Refineries	156	90	OGJ, Worldwide Refining Survey 1999
Iron & Steel	136	88	SteelEye, Survey 2001
Power Plants	2903	100	US EPA, E-Grid 2000
Cement	121	84	European Cembureau, World Directory 1996

**Characterization:** PH4/9 characterizes facilities with 46 fields for identification, location, CO<sub>2</sub> emissions, production, and fuels used. The report describes how many of the characteristics have been estimated based on the limited information available. A primary concern in considering the utility of the data is the accuracy of the CO<sub>2</sub> emissions numbers. The database

<sup>2</sup> <http://geonames.usgs.gov/gnishome.html>

has fields for both reported emissions and estimated emissions. The estimated emission is calculated from the plant's production, capacity, utilization factor and emission factor

**Timeliness:** The PH4/9 report was released in 2002, but the data in PH4/9 comes from a variety of sources published at different times. The data comes from information sources published between 1996 and 2001 (see Table 1).

**Modifications:** The spreadsheet is converted to CSV. The CSV file is then loaded into the Oracle database.

## ***CO<sub>2</sub> Sinks***

### ***Brine Formation Database***

The Bureau of Economic Geology at the University of Texas at Austin (UT-BEG) developed the *Brine-Formation Database* under a grant from the Department of Energy National Energy Technology Laboratory (DOE-NETL) in 2000.<sup>3</sup> UT-BEG designed the database to be GIS compatible and included power plant locations to match CO<sub>2</sub> emitters with brine formations. The researchers completed a detailed documentation of readily available data while completing the database.

UT-BEG selected 21 brine formations based on specific parameters that limited the scope of their exploration. The parameters, as listed in the documentation, were:

1. Geographic distribution of CO<sub>2</sub> sources;
2. Appropriate depth, injectivity, and seal;
3. Adequate information to characterize the target; and
4. Diverse geologic properties of the pool of selected formations.

The documentation includes the caution: "Many other brine formations besides those described in this study are suitable for use as sinks; the only reasons for not including them in this study were budget and time constraints."

UT-BEG evaluated the 21 brine formations for 16 characteristics. The characteristics included: top depth from surface, permeability/conductivity, formation thickness, net sand thickness, percent shale, sand-body continuity, top-seal thickness, continuity of top seal, hydrocarbon production from interval, fluid residence time, flow direction elevation, CO<sub>2</sub> solubility [temperature, pressure, salinity], rock/water reaction, porosity, water chemistry, and rock mineralogy. As discussed in the characterization section below, MIT is primarily interested in depth to the top of the formation, pressure, permeability, formation thickness, and seal integrity. The database does not contain all 16 characteristics for every selected formation. Despite the fact that UT-BEG used data availability as a formation selection criteria, there are a significant

---

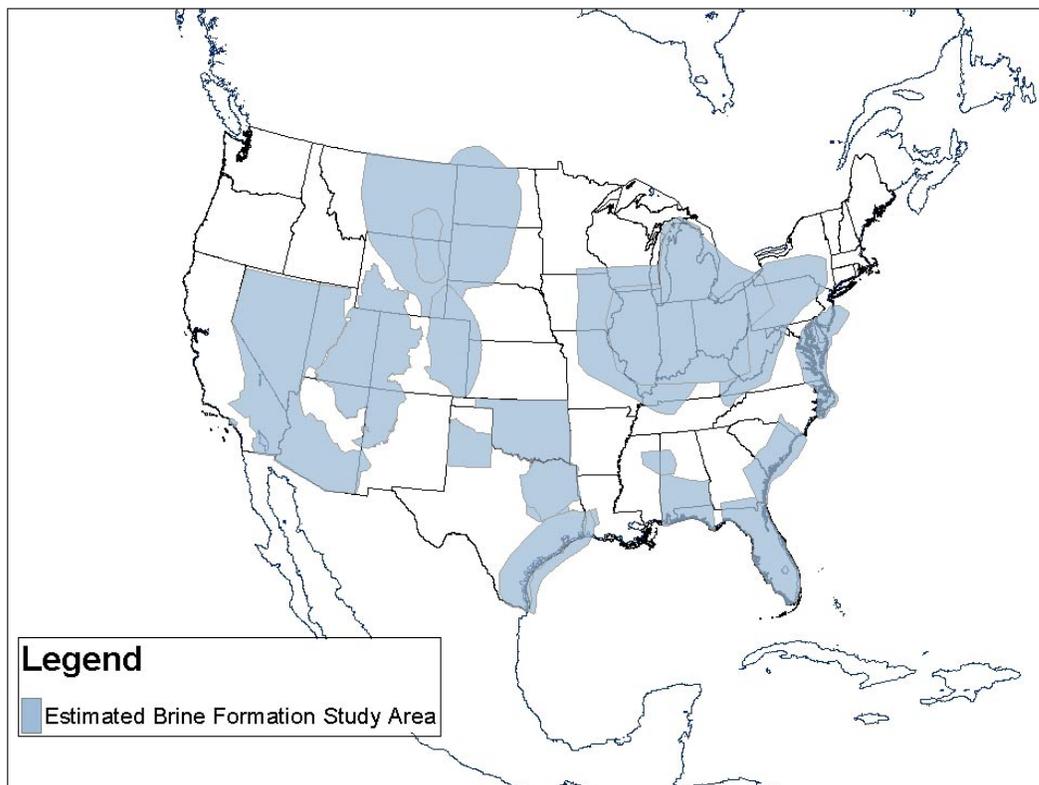
<sup>3</sup> Hovorka, Susan S. D., Martha M. L. Romero, Andrew A. G. Warne, William W. A. Ambrose, Thomas T. A. Tremblay, Ramon R. H. Treviño, and Douglas Sasson, 2000, Sequestration of Greenhouse greenhouse Gases in Brine Formations: <http://www.beg.utexas.edu/enviroqlty/co2seq/dispplsaln.htm>.

number of characteristics for which data does not exist. This reflects the fact that geologic investigators have not rigorously collected brine formation data in the past.

The remainder of this section includes comments on the usefulness of the data and notes on MIT's modification of the data for use in the MIT GIS.

**Usefulness of the data:** The UT-BEG database documentation encourages use for matching CO<sub>2</sub> sources with sinks; conducting preliminary feasibility studies; building economic and process models; and evaluating the merits of CO<sub>2</sub> reduction plans. MIT evaluated the data on the basis of coverage, characterization, and timeliness. MIT's evaluation concurs with the UT-BEG suggestion that the data are insufficient for detailed analysis. Significant improvements in data coverage and characterization are required to create a valuable economic and regulatory model. The sections below detail the data limitations.

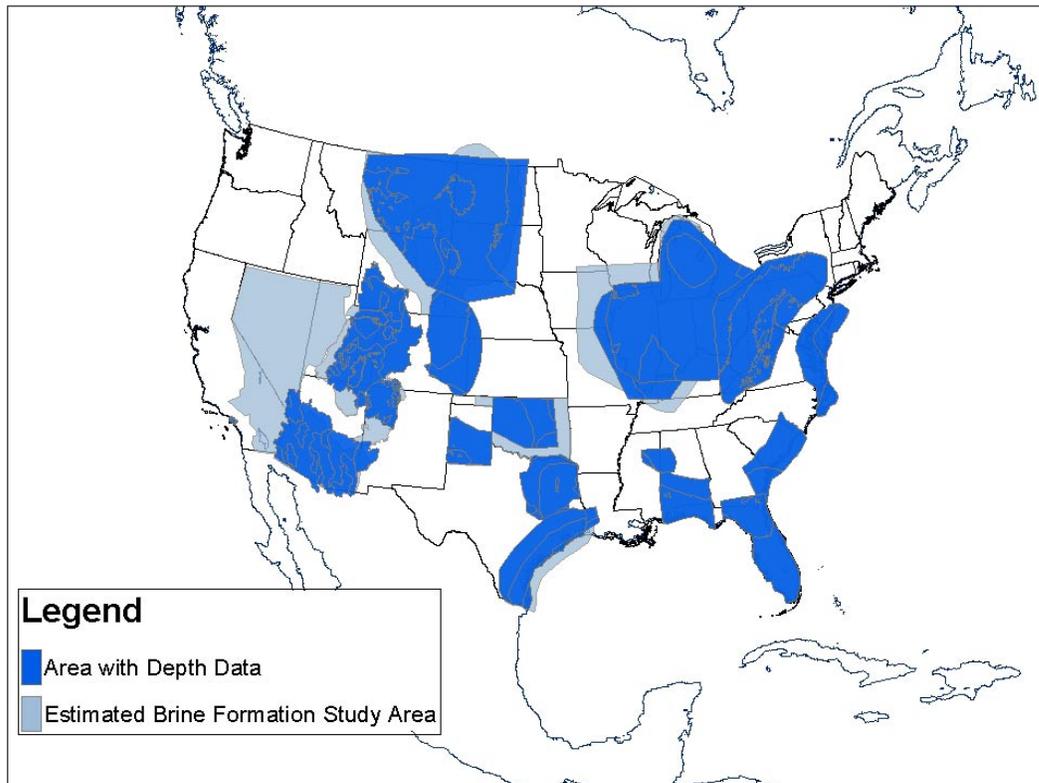
**Coverage:** The coverage of the database is limited. The parameters used by UT-BEG to choose the 21 formations in the database limit potential sinks. Figure 3 shows the estimated area covered by the formations investigated in the UT-BEG study.



**Figure 3: Estimated Brine Formation Study Area**

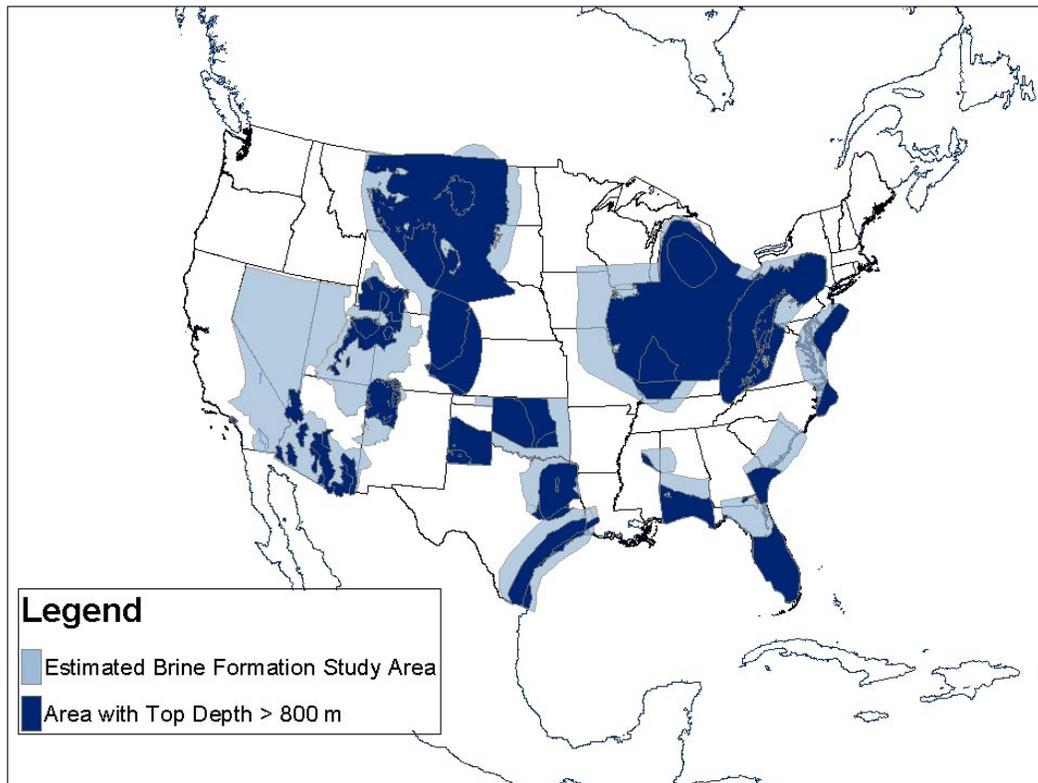
Figure 3 shows that there are areas of the country where brine formations are known to exist that did not meet the initial parameters established by UT-BEG. For example, the Midcontinent Interactive Digital Carbon Atlas and Relational dataBase (MIDCARB) contains detailed data on brine formations in Kansas. The UT-BEG study does not cover a large part of Kansas.

Additionally, coverage does not extend throughout the study area. Figure 4 shows the coverage provided by the most detailed set in the UT-BEG database, the formation depth data.



**Figure 4: Area with Depth Data**

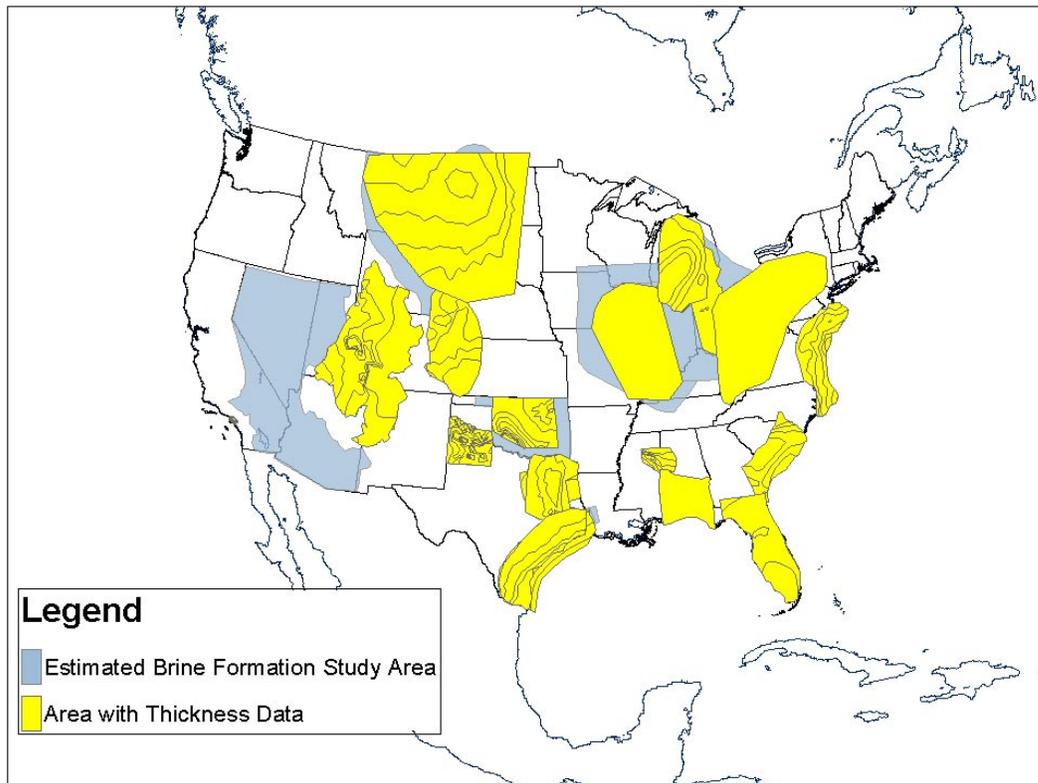
Figure 4 shows that there are significant areas of the study areas, especially in the west, that do not have depth data. The lack of coverage for some formations results in a shrinking of the area that a GIS can consider for injection. Depth is an important screening parameter because the injection depth must be greater than 800 meters to maintain adequate pressure on the injected CO<sub>2</sub>. Filtering the depth data shown in Figure 4, the GIS can show area where the depth to the top of the formation is greater than 800 meters. Figure 5 shows the results of this filtering.



**Figure 5: Area with Top Depth Greater than 800 Meters**

Comparing Figure 4 and Figure 5 shows reduced area available for injection especially in the western states and in the Gulf Coast. It is important to note that this reduction in coverage does not represent a limitation of the data; rather it represents the value of using the data for regional screening. Figure 4 shows the limits of coverage while Figure 5 shows the area within the limited coverage that researchers and industry should consider for injection.

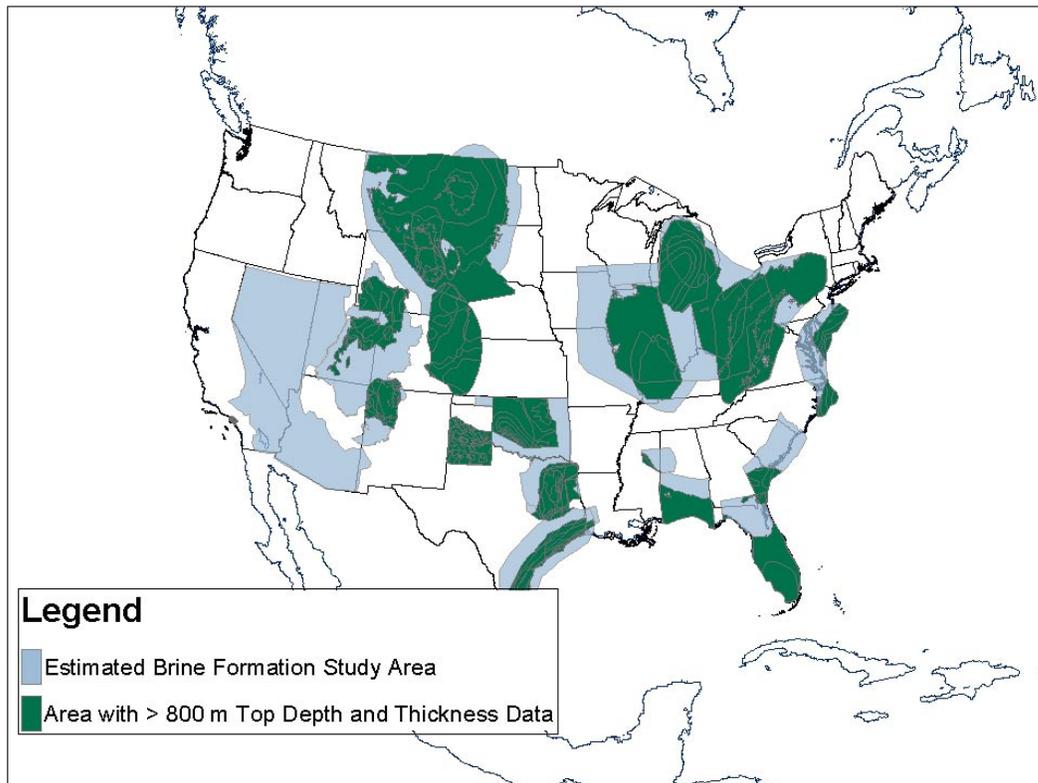
MIT also hopes to use formation thickness to find a cost for injection. Figure 6 shows the area of the country for which the UT-BEG database has formation thickness coverage.



**Figure 6: Area with Thickness Data**

Figure 6 shows that the UT-BEG database does not include thickness data in GIS-compatible form for California, Nevada, and most of Arizona. Additionally, Figure 6 reduced coverage in the area around the Great Lakes.

Combining Figure 5 and Figure 6 gives a picture of the area that can be considered for injection based on adequate depth and the existence of thickness data. Figure 7 shows the intersection of Figure 5 and Figure 6.



**Figure 7: Area with an Adequate Depth and with Thickness Data**

Figure 7 shows the area under consideration for CO<sub>2</sub> injection to be limited by the coverage of thickness and depth data. The UT-BEG database is especially lacking coverage in the western states (California, Nevada, and Arizona).

**Characterization:** MIT is primarily interested in depth to the top of the formation, pressure, permeability, formation thickness, and seal integrity. MIT can use these data along with distance between the CO<sub>2</sub> source and sink to estimate the cost of transportation and storage.

Table 2 provides a summary of the formations for which the database has data. An “x” signifies that there is data in the database for that formation. While most formations have data characterizing depth, permeability, and thickness, a lack of coverage limits the usefulness of the permeability data. There is no pressure data for seven of the formations. The pressure data that are characterized have limited coverage. Seal thickness and seal continuity data are each unavailable for four formations.

**Table 2: Availability of Characteristics for Brine Formations in the UT-BEG Database**

Formation	Depth	Permeability	Formation Thickness	Pressure	Seal Thickness	Seal Continuity
Arbuckle	x	x	x	x	x	x
Basin Fill/Carbonate	x	x				
Cape Fear	x	x	x	x	x	x
Cedar Keys	x	x	x	x	x	x
Fox Hills	x	x	x		x	x
Frio	x	x	x	x		x
Glen Canyon	x	x	x		x	x
Granite	x	x	x		x	x
Jasper	x	x	x	x	x	x
Lyons	x	x	x		x	x
Madison	x	x	x		x	x
Morrison	x	x	x		x	x
Mt. Simon	x	x	x	x	x	
Oriskany	x	x	x	x	x	x
Paluxy	x	x	x	x	x	x
Lower Potomac	x	x	x	x	x	x
Pottsville	x	x	x	x		x
Repetto	x	x	x	x	x	x
St. Peter	x	x	x	x	x	
Tuscaloosa	x	x	x	x	x	x
Woodbine	x	x	x	x	x	x

While there is good characterization of the formations, coverage limits the usefulness of the data. The pressure and permeability data is mostly in the form of point estimates that can not be easily extrapolated to larger areas. Pressure and permeability data that is available for larger areas of the formation frequently includes one estimate for a large area, for example the entire southern half of Florida has a permeability entered as 5-28 millidarcys. While that range can be used for initial screening, more detailed data are needed to perform a source to sink cost estimate.

**Timeliness:** UT-BEG collected the data in 2000. The database represents the most recent and detailed assessment of brine formations in the United States.

**Modifications:** Before importing the UT-BEG data into the MIT GIS, MIT manipulated and converted data rasters into shapefiles. Additionally, MIT filtered data rasters to find areas with adequate depth to the top of the formation. MIT perform similar manipulations to the thickness data. MIT studied pressure, permeability, and seal data in the downloaded form.

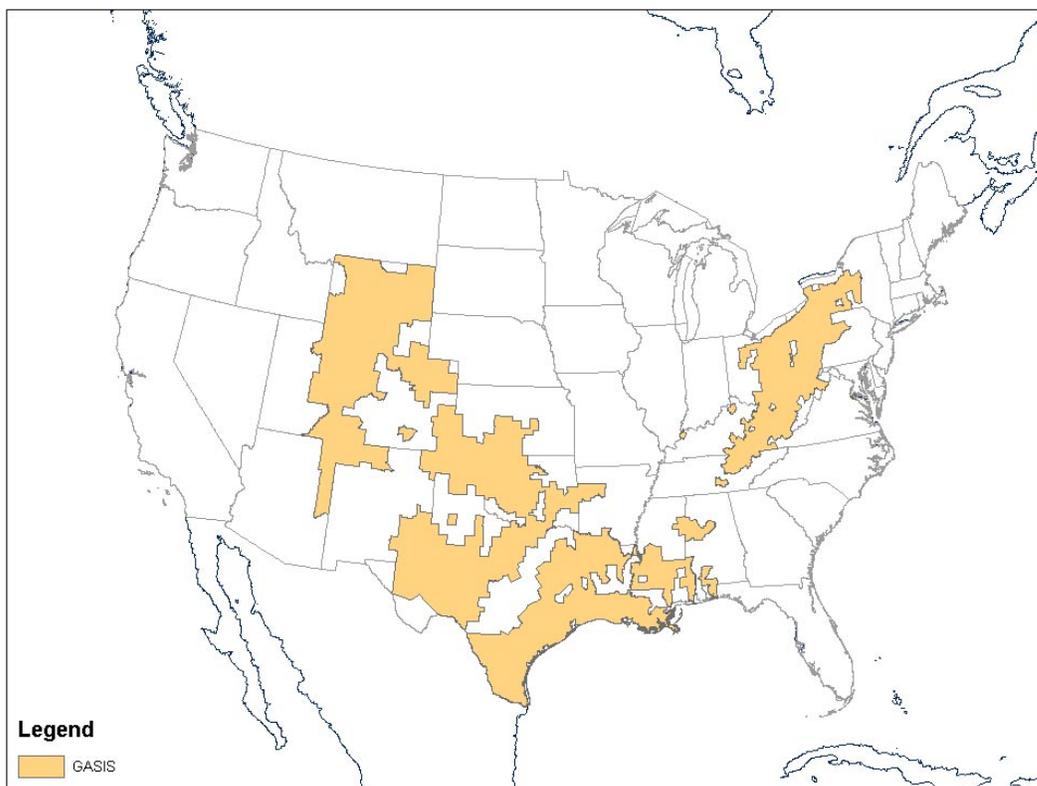
### **Gas Information System**

Energy and Environmental Analysis, Inc (EEA) developed the Gas Information System (GASIS) under contract with NETL between 1993 and 2000. The GASIS project was intended to provide a national database of gas reservoir properties and recovery data to aid natural gas producers in

field development and exploration. EEA merged data from the Department of Energy and Gas Research Institute's Gas Atlas data sets, Dwight's databases, and other public data sources.

**Coverage:** GASIS is a compilation of data sets such as the DOE gas atlas data sets with improvements to coverage and quality in a limited number of reservoirs. A major portion of the database consists of the Gas Atlas data sets for Texas, the Mid-Continent, the Central and Eastern Gulf Coast, the Rockies, the Appalachian Region, and the Gulf of Mexico. As such, the database coverage is limited to 21 states and the Gulf of Mexico.<sup>4</sup>

One issue encountered with the database is that coordinates are available for a limited number of gas fields (2633 of 19219), all in the Appalachian Region. However, the database includes county and state codes for all of the gas fields. Figure 8 shows the onshore coverage of the GASIS data.



**Figure 8: Onshore areas with reservoir data in GASIS**

**Characterization:** The GASIS database contains 185 data fields characterizing each gas field by identification, location, geological parameters, reservoir parameters, and production and reserve data. It includes data fields for depth, porosity, permeability, temperature, and pressure.

---

<sup>4</sup> SCNG: Projects – Development of a Gas Information System (GASIS)  
<http://www.netl.doe.gov/scng/projects/model/r-d/rdp28139.html>

**Timeliness:** Version 2 of GASIS was released in 1999. Production data in the database are available through 1996. Reservoir data contained in the database come from older databases such as the Gas Atlas. EEA improved reservoir property data on some reservoirs during the GASIS project.

**Modifications:** MIT obtained the GASIS data from NETL on a CD. The Dbase file on the CD was imported into the MIT database for use in the GIS system using an Oracle add-on.<sup>5</sup>

### ***National Oil and Gas Assessment***

The National Oil and Gas Assessment is an ongoing project by the USGS with the purpose of assessing oil and gas reserves of the US and determining the ability of these reserves to maintain production capacity in the future. As described in the summary:

The approach is to 1) prioritize areas and commodities to be assessed, 2) define total petroleum systems in priority provinces, 3) define assessment plays within these petroleum systems, 4) assess the plays using established methodologies, and 5) forecast the availability of undiscovered resources, including resources under Federal lands.<sup>6</sup>

The USGS provides data from the 1995 assessment to the public in the form of internet downloads or CDs. The distributed data includes national data on wells and plays.

The national well data is provided on the resolution of quarter-mile cells or one-mile cells. In either case, each cell is labeled with the number of wells in the cell and as oil producing, gas producing, oil and gas producing, or dry. No more information is provided because the data comes from the proprietary Petroleum Information Well History Control System (WHCS) of 1990.<sup>7</sup>

A play is “defined as a set of known or postulated oil and or gas accumulations sharing similar geologic, geographic, and temporal properties, such as source rock, migration pathways, timing, trapping mechanism, and hydrocarbon type.”<sup>8</sup> For each play, oil, associated gas, associated gas liquids, non-associated gas, and non-associated gas liquids are estimated. The estimation is represented by a mean quantity and a histogram.

MIT has chosen not to focus effort on this data because neither the well data nor the play data provide reservoir properties. Properties such as porosity and permeability are important to an accurate prototype, and are in the GASIS database.

---

<sup>5</sup> <http://asktom.oracle.com/pls/ask/f?p=4950:8:326259121521467429>

<sup>6</sup> Shenk, Christopher. *National Assessment of Oil and Gas Resources*. USGS, Denver.  
<http://certmapper.cr.usgs.gov/data/noga95/natl/text/summary.pdf>

<sup>7</sup> National Oil and Gas Assessment – one-quarter mile cell data: Metadata  
<http://certmapper.cr.usgs.gov/data/noga95/natl/spatial/doc/usqtmicg.htm>

<sup>8</sup> 1995 National Oil and Gas Assessment Plays: Metadata  
<http://certmapper.cr.usgs.gov/data/noga95/natl/spatial/doc/nogaplays.htm>

### ***Total Oil Recovery Information System (TORIS)***

The Total Oil Recovery Information System (TORIS) – alternately known as the Tertiary Oil Recovery Information System – was originally developed by the National Petroleum Council (NPC) in 1984 for the DOE in an assessment of EOR potential. It has been maintained and updated by the National Petroleum Technology Office (NPTO), but only the data from the 1984 assessment is available to the public. The system consists of technical data for oil fields and reservoirs as well as engineering models.

The TORIS Reservoir/Geologic data file contains 61 data fields for over 2500 oil reservoirs. The data file includes important factors such as porosity, permeability, temperature and depth of the reservoirs. However, the field location is not included, only a field code. Because of the difficulty in locating the fields, MIT is not integrating the TORIS data at this time.

### ***Coal Fields of the United States***

The primary dataset used in MIT’s GIS for coal sinks is the “Coal Fields of the United States” published by the United States Geological Survey (USGS). This dataset is the most complete and useful for analyzing potential sinks for CO<sub>2</sub> storage in coal fields in the United States. It is a polygon coverage representing all coal fields in the US.

While this dataset provides useful information for a broad overview of US CO<sub>2</sub> storage options, it does not contain enough information for evaluating and selecting individual project sites. The data resolution is not high enough (resolution is 1:5,000,000), nor is there enough data on requisite parameters such as porosity, permeability, and other information needed to make confident planning decisions.

**Coverage:** This dataset covers all major coal deposits located within the 50 US states, including the District of Columbia.

**Characterization:** The dataset characterizes coal deposits based on the Geologic Names Unit Lexicon. Coal types are broken down into the following categories:

- Lignite
- Sub-bituminous
- Bituminous
- Anthracite
- Semianthracite
- Meta-anthracite

Each coal deposit has also been characterized as either suitable for mining, or unsuitable for mining. The metadata suggests that the unminable seams may be suitable for carbon sequestration. However, some of these sites characterized for sequestration, may be suitable for mining, but not enough is currently known about them.

**Timeliness:** The majority of this dataset was collected from the USGS “Coal Fields of the United States” from 1960. The data from Alaska was collected from USGS “Coal Fields of the

United States, Alaska” produced in 1961. Some data from the Gulf Coast region is more recent, being collected from various state maps. Information on current existing coal resources is contained in the USGS NEWCOAL database. NEWCOAL is not currently accessible to the public.

**Limitations:** The dataset is limited in its ability to offer a robust policy analysis for a GIS decision support system. Information on porosity, permeability, depth, seam thickness, composition of overlying cap rock, and down-hole pressure are all unknown. This means that assumptions for all these variables must be made to calculate the volume of the CO<sub>2</sub> that can be stored, and to calculate injectivity (basically, how difficult is it to pump CO<sub>2</sub> down the hole), which directly impacts the economics of storage.

Another limitation in this field is the uncertainty of the geo-chemical/physical processes that occur when CO<sub>2</sub> is injected into a coal seam. It is postulated that as CO<sub>2</sub> is injected, the micropores in the coal will become clogged as CO<sub>2</sub> attaches itself to the coal. However, the rate at which this will happen, or what it will mean for overall reservoir capacity, or the economics of injection are unknown. Also, different grades of coal are expected to have affinities for CO<sub>2</sub>, but the magnitudes of the differences are unknown.

**Modifications:** We filtered the database to create a new layer for all coal sites that are classified as potential CO<sub>2</sub> storage sites. This layer can be used as a separate file, and joined with information from other database to add pertinent information.

## Conclusion

Work has progressed according to schedule in year one of this project. Our plan is to continue this project in line with our original proposal. Specific milestones for year 2 include include:

- Complete Task 1 (detailed system design) and issue the task 1 report.
- Complete Task 2 (create core data sets) and issue the task 2 report.
- Complete working prototype (GIS for US including major sources and sinks)
- Development of initial data analysis tools (task 4) and their incorporation into the prototype system.

Work on both task 3 (incorporate economic, regulatory, and political data) and task 4 (develop analysis tools) will also commence in year 2. In addition, we are members of a DOE working group which will develop a National Atlas for key GIS data.

## References

Heddle, G., H. Herzog and M. Klett, “The Economics of CO<sub>2</sub> Storage,” MIT LFEE 2003-003 RP, August (2003).