GPS/GIS FOR REGULATED FACILITIES IN THE EPA REGION 6 INTERNATIONAL BORDER AREA

by

E. S. Angle, G. White, D. Salazar, A. Goldberg, and J. Raney

Final report prepared for EPA Region 6 under contract no. X-996361-01-2

Bureau of Economic Geology
Noel Tyler, Director
The University of Texas at Austin
Austin, Texas 78713-8924

September 1996

GPS/GIS FOR REGULATED FACILITIES IN THE EPA REGION 6 INTERNATIONAL BORDER AREA

by

E. S. Angle, G. White, D. Salazar, A. Goldberg, and J. Raney

Final report prepared for EPA Region 6 under contract no. X-996361-01-0

Bureau of Economic Geology
Noel Tyler, Director
The University of Texas at Austin
Austin, Texas 78713-8924

September 1996

CONTENTS

ABSTRACT 1

INTRODUCTION 1

METHODS 3

Data Organization 3

Fieldwork 5

Equipment 6

RESULTS 8

DGPS data 8

ArcInfo Coverages 9

DISCUSSION 9

Project Evolution 9

Cost Analysis 11

CONCLUSION 12

ACKNOWLEDGMENTS 13

APPENDIX A 14

Figures

1. Map showing EPA-GIS field area 4
2. Bull's-eye plot comparing NGS reference point to DGPS file 7

Tables

1. Data bases for the EPA-GIS Project 2
2. Site-visit data 8
3. Record total by county 9
4. Trip efficiency 11
5. Direct costs of DGPS data collection 12

ABSTRACT

The University of Texas at Austin, Bureau of Economic Geology (Bureau), conducted a 2-yr project for the Environmental Protection Agency (EPA), Region 6, using real-time Global Positioning System (DGPS) technology to locate regulated facilities in Texas and New Mexico along the Mexico border. Accurate locations of these facilities are required to establish a Geographical Information System (GIS) that will provide State and Federal regulators with a more functional tool for their oversight responsibilities. The Texas Natural Resources Conservation Commission (TNRCC) and the New Mexico Environmental Department (NMED), along with EPA Region 6, provided records from their regulatory data bases that include toxic release inventory, industrial waste producers/transporters, Superfund sites/landfills, and underground storage tanks. These facilities are described in 7,380 records within 10 data bases. The goal of this project is to provide spatial data in ESRI's ArcInfo coverages on 1,000 to 1,500 facilities. Using a Pathfinder Basic Plus and an Omnistar receiver, a single fieldworker can collect as many as 18 sites per day in urban settings such as the city of El Paso, Texas, or about 7 sites per day in more rural areas such as Cameron County, Texas. To date, we have collected locational data on 1,116 facilities from 6 Texas counties and 1 New Mexico county that relate to 2,371 records across 10 regulatory data bases. The coordinates for these facilities have been loaded into ArcInfo coverages. Efficient application of this technology has resulted in a much improved set of spatial data that can be used to link other records on these regulated facilities to actual locations portrayed in a GIS.

INTRODUCTION

In support of the EPA Locational Data Policy (LDP) instituted in May 1990, the Bureau explored the use of real-time, differentially corrected Global Positioning System technology (DGPS) to develop a method for centralized geocoding. The primary focus of this project is the collection of horizontal and vertical locational data on 1,000 to 1,500 facilities along the

U.S./Mexico international border that are regulated and permitted by the EPA, Region 6, TNRCC, and NMED. In the past, GPS data collection presented problems because the technique was expensive and required postprocessing of the field data, a complicated and time-consuming procedure. DGPS eliminates the need for postprocessing and allows a corrected position to be collected in real time. Although a number of DGPS systems exist, we selected the Omnistar system because it offered complete coverage of the field area and could be used for future projects at the BEG.

The original project parameters for data collection stipulated that the accuracy of the locations at each facility should be less than 25 m horizontally and vertically (95-percent probability). At each facility, a point would be located to represent the street address, preferably within 25 m of the entrance to the facility. At facilities greater than 250,000 ft^2, points along the boundary would be recorded, if readily accessible, so that a polygon could be drawn to represent the facility. Because of the scope of the project, we decided not to obtain private access to a facility or to determine legal boundaries. Although this decision limited our access of facilities to public roads and walkways, at each facility the field technician attempted to contact responsible parties before collecting GPS data. This courtesy call was made to inform the proprietors of our presence and intent.

EPA, NMED, and TNRCC provided the Bureau with 10 regulatory data bases, as listed in table 1, containing a total of 7,380 records.

Table 1. Data bases for the EPA-GIS Project.

FINDS—Facility Index System (Master File for PCS, CERCLIS, TRIS, RCRIS)
CERCLIS—Comprehensive Environmental Response, Compensation, Liability Information System (Superfund Program)
TRIS—Toxic Release Inventory System
RCRIS—Resource Conservation and Recovery Information System (Hazardous Waste)
AIRS—Aerometric Information Retrieval System (Air Pollution Data)
PCS—Permit Compliance System (Water Discharge)
FRDS—Federal Reporting Data System (Drinking Water)
FTTS—Federal Toxic Tracking System (PCB's, FIFRA)
NMED—New Mexico Environmental Department (Water, Underground Storage Tanks)
TNRCC—Texas Natural Conservation Commission (Landfill, Wastewater, Hazardous Waste Producers, Superfund Data)

These data bases cover the U.S./Mexico international border from Cameron County in South Texas through Hidalgo County in southwestern New Mexico. Because the field area is so aereally extensive that all facilities could not be visited, counties containing the largest number of records were chosen as primary locations for data-collection trips. Willacy, Cameron, Hidalgo, Webb, Val Verde, and El Paso Counties in Texas and Doña Ana County in New Mexico were selected as primary areas for fieldwork. Except for Willacy County, these counties represent the most populated counties along the border (fig. 1).

METHODS

This project involved two principal tasks: (1) organizing the various data bases provided by the regulating agencies and (2) collecting locational data in the field. Using Microsoft Access, we compiled the original 10 data bases that contained all facility records into a single data base called EPATRACK. The locational data were placed in coverages created in ESRI's ArcInfo, with corresponding facility information as attribute tables.

Data Organization

The original tracking data bases contained various formats and data fields. It was necessary to combine them into one data base to ensure that no data would be missed during mission planning for fieldwork. Our main focus when combining these original data bases was to preserve the integrity of each data set and each individual record. The 10 data bases were first assembled together on the VAX computer and then exported to Microsoft Excel via FTP. In Excel, a tracking number was assigned to each record, and the entire file was exported to Microsoft Access into a data base called EPATRACK. The EPATRACK data base was designed so that facility data could be printed to field sheets to be taken to the field and updated there. Microsoft Access is a relational data base software package that runs under Windows. Providing a user-friendly interface that is easily customized, it has an object-oriented form of SQL (Structured Query Language) that allows

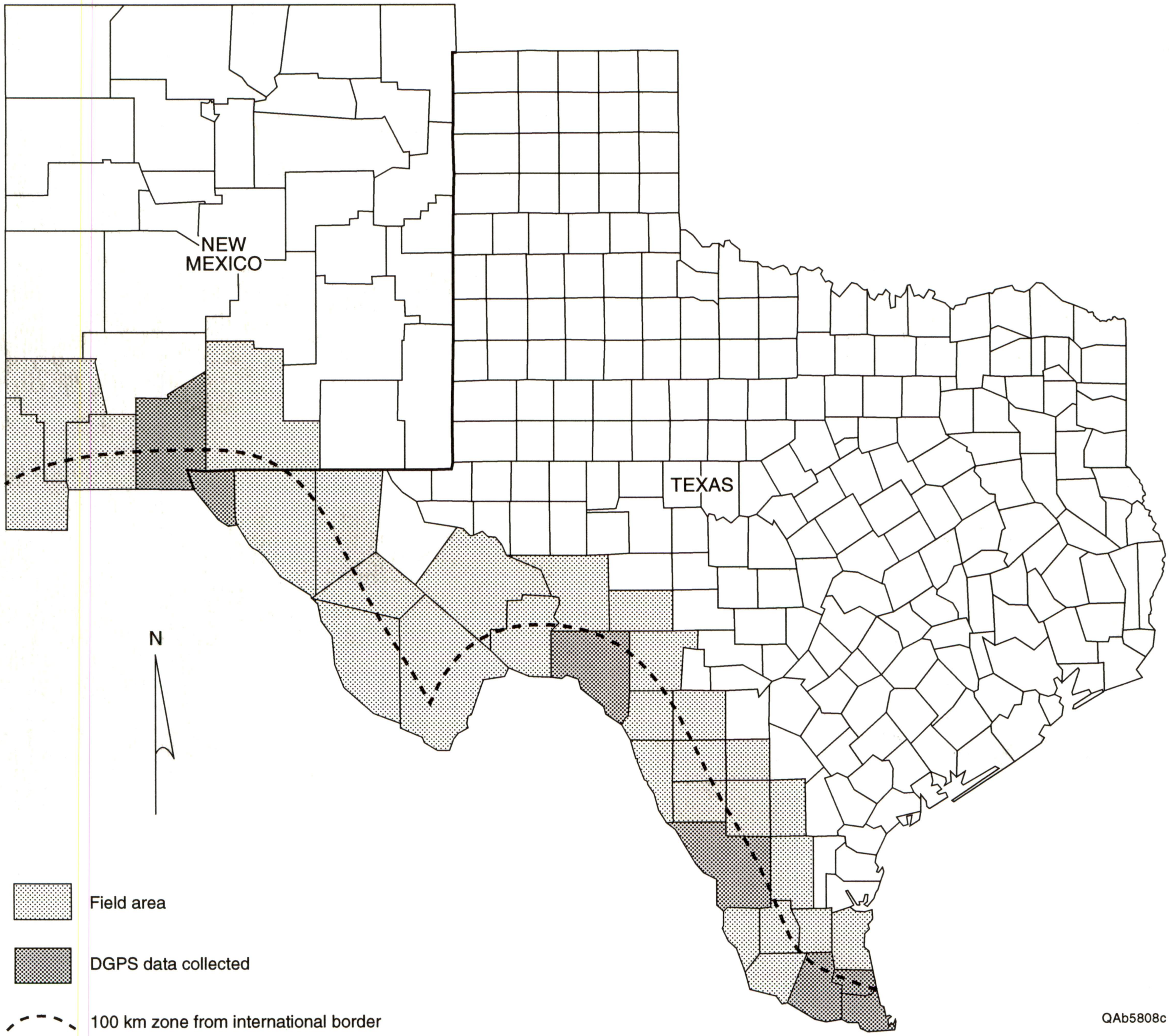


Figure 1. Map showing EPA-GIS field area.

the user to design macros and queries inside the data base. A series of macros and queries were written to manage the data base for the life of the project. The EPATRACK data base consists of 7,380 rows, each containing a single record, and 36 columns for all the information required to locate and track each facility. Because the column information varied among the original data bases, some columns in the EPATRACK data base are not always complete for every record.

Within the original 10 data bases, a large number of incomplete and duplicate records translated across to EPATRACK in Microsoft Access. Any record that was missing the facility name or address was considered incomplete, and these records were culled later during mission planning. Records that contained identical data in the facility name, address, and source identification number columns were considered duplicates. By using the unique identification number assigned by the BEG, we tracked duplicate records and linked them to one record chosen as the master reference. A Microsoft Access query was written to update fields across the data base when any field in the master reference was changed.

Fieldwork

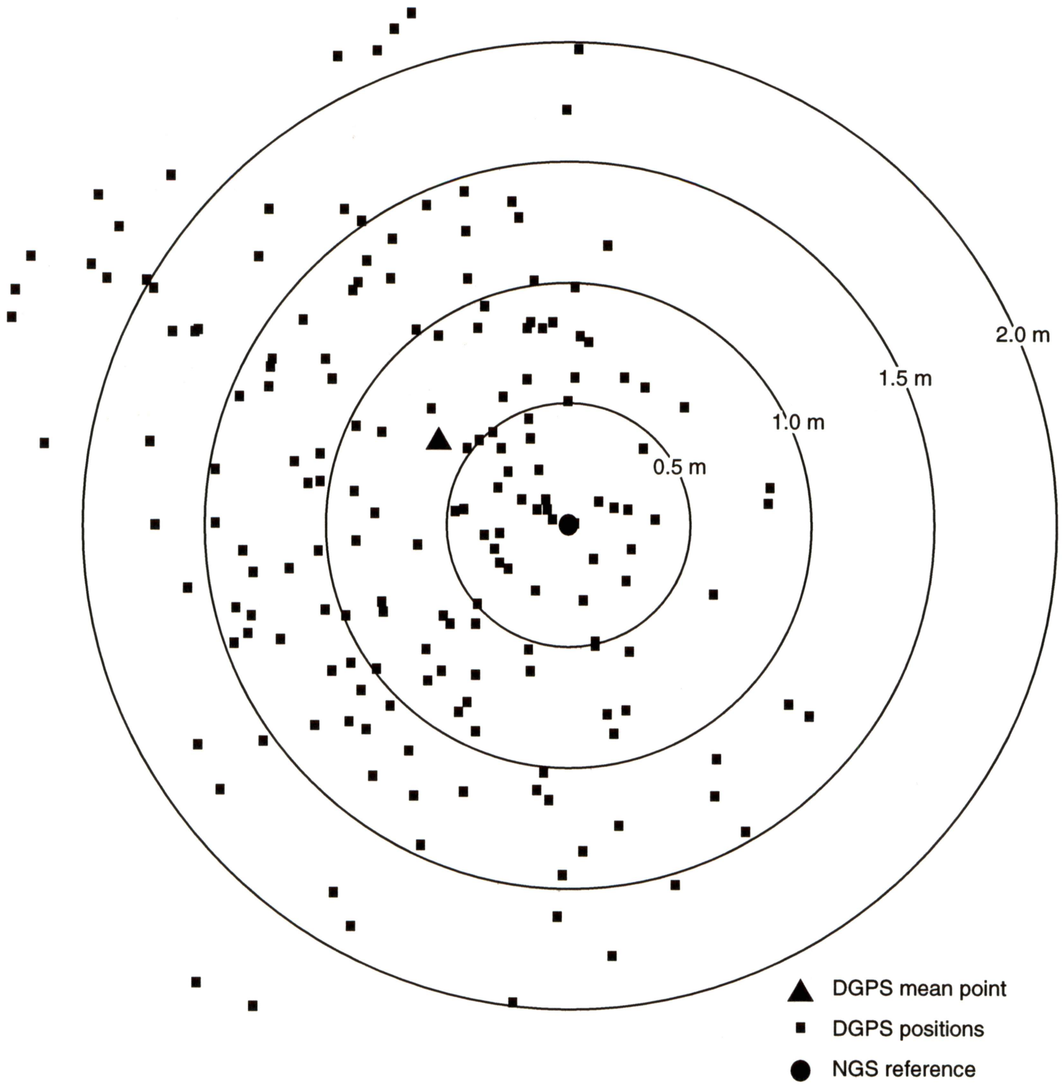
Planning and organization are key to efficient data collection in the field. Because of the large field area and number of sites, careful planning of field trips was essential. Each trip consisted of two phases: mission planning and field-data collection. During mission planning, we organized the data for each trip by county. We queried EPATRACK for all data in a given county using the FIPS code, and these records were written to a separate table within Microsoft Access. Each record was reviewed to ensure that the address and facility name were complete. Many records contained incomplete addresses, often missing street numbers, or addresses were listed only as post office box numbers or verbal descriptions (for example, "6 miles east Hwy. 123"). These records were put aside to be researched at a later date. The remaining records were printed on field-data sheets. The records were then divided up into their respective cities and further grouped by street address. Using either a county road map or the corresponding city road map, we color coded each record to the map by street name. Each street was assigned a color, and records containing that street address were coded in the same color. All records were then organized by city and street into a binder to be taken to the field, where an attempt was made to locate each facility listed on the field sheets.

Equipment

We collected the DGPS data using a Trimble Navigation Pathfinder Basic Plus unit. This receiver is a hand-held, battery-powered, six-channel receiver. It can track as many as eight satellites simultaneously and has 256 Kbytes of nonvolatile memory. Positions can be calculated at a rate of one per second and stored for later transfer to a personal computer for processing on Trimble software, P-FINDER. The accuracy of the Pathfinder is rated by Trimble at 3 to 5 m horizontally, on the basis of the average of 180 data points from a differentially corrected file.

We achieved real-time differential correction using an Omnistar Model 6300A receiver. The Omnistar system is based on 11 base stations in North America that monitor and send corrections to a network control center in Houston, Texas, where the data are uplinked to a geostationary satellite. The satellite then broadcasts the corrections to clients having Omnistar receivers. The accuracy of the Omnistar is rated at ±1 m. Reference data collected throughout the project at first-order NGS survey points support these claims (fig. 2).

The Pathfinder and Omnistar receivers were carried in a backpack padded with foam to minimize sharp impacts, but the added insulation also reduced heat dissipation and allowed the Omnistar to overheat on one occasion. To prevent overheating, a cardboard insert was made to hold the Trimble and Omnistar receivers in place so that air could flow around them and keep the units in place. The system worked well and was maintained for the life of the project.



Trimble DGPS file	Mean coordinates	Standard deviation (m)
Latitude	31 48 12.197	0.861
Longitude	106 17 41.555	0.671
Altitude	1224.058 m	2.617

NGS reference CE 0401	Mean coordinates
Latitude	31 48 12.187
Longitude	106 17 41.533
Altitude	1223.390 m

Datum: NAD-83
Coordinate system: latitude/longitude
altitude mode: height above ellipsoid

QAb5807c

Figure 2. Bull's-eye plot comparing NGS reference point to DGPS file.

RESULTS

DGPS data

Over the course of the project, 1,427 sites were visited, as shown in table 2. Of these, DGPS location data were collected at 1,116 sites. Another 311 sites were visited at which the facility referenced in the record was absent.

Table 2. Site-visit data.

County name	Sites visited (with data)	(without data)	Total
Cameron	207	83	290
Hidalgo	119	53	172
Willacy	12	0	12
Webb	61	0	61
Val Verde	60	0	60
El Paso	534	148	682
Doña Ana	123	27	150
Total	1,116	311	1,427

Records in the ACCESS data base can be separated into three categories: (1) records of sites that have GPS locational data (table 2, Sites visited with data), (2) records of sites visited but where no locational data were collected (table 2, Sites visited without data), and (3) records of sites that are duplicates linked to the master reference record. Category 1 records represent sites that we visited and at which we collected DGPS location data, either as a single point or as a polygon. Records in category 2 represent sites where we attempted a visit but the facility in question was not found at the given address. In many cases, addresses led to sites where no facility existed, a different facility existed, or the site consisted of an abandoned building or open field. If possible, we queried the current occupants about the location of the facility in question. This information was logged and later added to the "comments" section in the Microsoft Access data base. Three hundred eleven records had no locational data collected. In category 3, these records are duplicates that are linked to master records and updated when the master records are located. At the completion of the project, 2,371 of the 4,975 records located in the 7 counties visited were updated with DGPS location data, as shown in table 3.

Table 3. Record total by county.

County	Total number of records	Records with locational data	Completed (%)
Cameron	797	444	56
Hidalgo (TX)	426	206	48
Willacy	77	23	30
Webb	319	123	38
Val Verde	169	121	71
El Paso	2,344	1,198	51
Doña Ana	842	256	30
Total	4,974	2,371	46

Arcinfo Coverages

The P-FINDER software package allows GPS data to be converted into the ArcInfo format, which can then be exported to ArcInfo. Coverages in ArcInfo were organized according to the county in which data were collected. Each county has two coverages: one with sites larger than 250,000 ft^2 that are represented by polygons and the other with sites smaller than 250,000 ft^2 represented by single points. Sites in the coverages each have a reference in an ArcInfo file that contains all EPATRACK data associated with that site. The coverages are single precision and projected in Universal Transverse Mercator, with horizontal and vertical measurements in meters. North American Datum 1983 is used as the vertical reference datum.

DISCUSSION

Project Evolution

Cameron County was chosen as the test area in which to refine data-base construction ideas and data-collection methods and to determine the feasibility of using a single field technician. The first trip resulted in 130 site visits and collection of locational data on 73 sites over a 10-day period.

The initial parameters of the project required that reference data be collected at a local first-order National Geodetic Survey (NGS) site in the field twice each day to verify the accuracy of the DGPS data collected during that trip. The daily routine was structured such that each morning NGS reference data were collected, facility data were collected during the day, and at the end of the day a return trip was made to the NGS site. Because most of the first-order NGS makers in Cameron County have been destroyed or can no longer be located, approximately 1.5 to 2 h were spent each day visiting the NGS site. This also proved to be true in later visits to other counties. The data collected at NGS sites consistently fell within 1 to 3 m of the reference point. To save time during future trips, the number of NGS references was reduced to two, one at the beginning and one at the end of the trip.

During the Cameron County trip we learned that it would be feasible for one field technician to collect data for the project. Collecting DGPS data in the field was divided into two major tasks: getting to the site and collecting the data. It became apparent early on that travel time to sites was dictated by site density. In highly urban areas such as El Paso, as many as 30 sites a day could be visited. In Webb County's more rural setting, 10 to 14 site visits were more the norm. Two tasks were performed at the site: making a courtesy call to inform someone at the facility of our presence and collecting the DGPS data. At first, much time was spent making courtesy calls at each site. The field technician would enter the facility to inform the occupants that we would be collecting data near their facility. About half the time it was not a problem, but often the secretary or the first contact person would want to refer the field technician to a manager or someone higher up, and much time would be lost waiting to speak with this person. On all trips during the 1995 season, a courtesy call was made at all sites; later, during the 1996 field season, this practice was followed only when necessary. The actual collecting of DGPS data was not very time consuming. The Pathfinder will collect 40 to 45 data points per minute. Each file of 180 data points takes anywhere from 4 to 6 minutes to log.

Other problems were encountered in urban areas. The City of El Paso posed some unique problems because of its size and high concentration of sites. It became more difficult and time consuming to collect multiple points for polygon data at large sites. Accessibility to large facilities is more limited in highly urbanized areas, and often no place exists at which to record data without entering the facility. To streamline data collection and save time, we suspended collecting polygon data for the 1996 field season.

In all, we were able to average about 14.7 sites per day during the 100 (actual) field days. In table 4, site visit per day by county is shown in the "Average sites/day" column. High visit rates occurred in counties having large cities, such as El Paso County (El Paso), Hidalgo County (McAllen), and Doña Ana County (Las Cruces). The greater efficiencies, as measured in sites per day, occur because facilities are often concentrated in industrial parks or particular sections of the city, and travel times are reduced between sites. In counties such as Cameron County, many facilities were spread out in more rural, agricultural areas, and travel time between sites was increased.

Table 4. Trip efficiency.

County	Number of trips	Working days	Total site visits	Average sites/day
Cameron	4	27	290	10.7
Hidalgo (TX)	2	10	172	17.2
Willacy	1	1	12	12
Webb	1	9	61	6.7
Val Verde	1	8	60	7.5
El Paso	5	37	682	18.4
Doña Ana	2	8	150	18.7
Total	13*	100	1,427	14.3

*Trips to Doña Ana County, New Mexico, were combined with El Paso County, Texas, trips because of proximity.

Cost Analysis

In an attempt to assess the actual cost of this method of geocoding, we examined the direct cost of putting someone in the field to collect DGPS data. Table 5 is based on the following assumptions: (1) Training on software and DGPS equipment was included in labor costs because no external sources were used. (2) Data-base development time and field-equipment design were also included in labor costs because no external sources were used. (3) Per diem and hotel costs were based on rates approved by the State and University. (4) Transportation costs were not itemized; they are based on average rates. (5) Vacation and benefits are included in salaries for labor. (6) No overhead was included in this assessment.

Table 5. Direct costs of DGPS data collection.

			Cost
Labor			
	1 field technician	$136/day × 359 days = $48,824	
	1 assistant	$50/day × 84 days = $4,200	
			Subtotal $53,024
Field			
	Meals	$25/day × 123 =$3,075	
	Vehicle	$50/day × 123 =$6,150	
	Airfare	$250/trip × 5 =$1,250	
	Lodging	$50/day × 123 =$6,150	
			Subtotal $16,625
Equipment			
	GPS equipment		
	DGPS receiver	$3,000	
	GPS receiver	$6,000	
	DGPS subscription	$3,000	
	Miscellaneous	$500	
	Other		
	Laptop computer	$2,800	
			Subtotal$15,300
			Total $84,949
Cost per position			
	$84,949/1,116 GPS positions = $76.12 per position		

CONCLUSION

The current goal of EPA Region 6 is to have locational data for all its regulated facilities throughout the five-state region by the end of calendar year 2000. In its 1990 Locational Data Policy, the EPA recognized GPS technology as being the most accurate method for providing locational data to update its data bases. The use of real-time DGPS technology enabled this project to provide locational data on 1,116 facilities in 100 days of fieldwork for 7 border counties in

Texas and New Mexico. The use of this technology will help ensure that EPA Region 6 meets its goal.

ACKNOWLEDGMENTS

Funding for this project was provided by EPA Region 6 under contract no. X996361-01-0. Logistical support for fieldwork, data-base conversion, and ArcInfo was provided by BEG employees Allison Goldberg, Gerry White, and Erika Boghici. Word processing and pasteup were by Susan Lloyd. Lana Dieterich edited the report. Publication authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

APPENDIX A

The following tables show records with and without locational data, first in table 6 for the total counties visited and then in tables 7–13 by individual county.

Table 6. Record total for all counties visited by data base.

Data base	Total number of records	Records with locational data
AIRS	159	62
CERCLIS	181	73
FINDS	597	457
FRDS	991	237
FTTS	151	83
PCS	192	69
RCRIS	1308	766
TNRCC	1284	534
TRIS	48	38
NMUST	64	52
Total	4975	2371

Table 7. Record total by data base for Willacy County.

Data base	Total number of records	Records with locational data
AIRS	3	0
CERCLIS	9	5
FINDS	3	3
FRDS	41	7
FTTS	0	0
PCS	5	2
RCRIS	15	6
TNRCC	0	0
TRIS	0	0
Total	76	23

Table 8. Record total by data base for Cameron County.

Data base	Total number of records	Records with locational data
AIRS	24	16
CERCLIS	46	37
FINDS	98	85
FRDS	56	39
FTTS	29	21
PCS	80	21
RCRIS	185	124
TNRCC	268	92
TRIS	11	9
Total	797	444

Table 9. Record total by data base for Hidalgo County.

Data base	Total number of records	Records with locational data
AIRS	14	3
CERCLIS	60	12
FINDS	89	63
FRDS	3	0
FTTS	14	1
PCS	42	17
RCRIS	199	107
TNRCC	1	1
TRIS	6	2
Total	428	206

Table 10. Record total by data base for Webb County.

Data base	Total number of records	Records with locational data
AIRS	19	4
CERCLIS	7	3
FINDS	39	31
FRDS	22	0
FTTS	3	3
PCS	6	0
RCRIS	100	53
TNRCC	122	29
TRIS	1	0
Total	319	123

Table 11. Record total by data base for Val Verde County.

Data base	Total number of records	Records with locational data
AIRS	0	0
CERCLIS	3	1
FINDS	10	10
FRDS	65	53
FTTS	5	4
PCS	6	6
RCRIS	37	29
TNRCC	41	16
TRIS	2	2
Total	169	121

Table 12. Record total by data base for El Paso County.

Data base	Total number of records	Records with locational data
AIRS	38	23
CERCLIS	38	15
FINDS	317	241
FRDS	276	34
FTTS	84	46
PCS	20	6
RCRIS	692	412
TNRCC	852	396
TRIS	27	25
Total	2344	1198

Table 13. Record total by data base for Doña Ana County.

Data base	Total number of records	Records with locational data
AIRS	61	16
CERCLIS	18	0
FINDS	41	24
FRDS	528	104
FTTS	16	8
NMUST	64	52
PCS	33	17
RCRIS	80	35
TRIS	1	0
Total	842	256