UNUSUALLY SENSITIVE AREAS

PROJECT GIS TECHNICAL REVIEW

For the

U.S. Department of Transportation
Research and Special Programs Administration
Office of Pipeline Safety

May 19, 2000

Baker

Michael Baker Jr. Inc.
A Unit of Michael Baker Corporation

3601 Eisenhower Ave., Suite 600
Alexandria, Virginia 22304

(703) 960-8800
FAX (703) 317-3079

UNUSUALLY SENSITIVE AREAS PROJECT GIS TECHNICAL REVIEW

Introduction

Michael Baker Jr. Inc. (Baker) was asked by the U.S. Department of Transportation (USDOT), the Research and Special Programs Administration (RSPA), Office of Pipeline Safety (OPS) to provide a technical review of its effort to identify geographic areas unusually sensitive to damage from hazardous liquid pipeline release. These areas are to be known as USAs. OPS, working in cooperation with the American Petroleum Institute (API), has completed a pilot study in which models of two types of USAs were defined: the drinking water USA model and the ecological USA model. As part of a pilot study, the two models were tested on data from three states-- California, Louisiana, and Texas.

Baker is one of several technical/peer reviewers asked to comment on the USA models and how they were applied during the pilot study. However, unlike other reviewers, Baker was asked to pay particular attention to how geographic information system technology (GIS) was applied during the pilot study. GIS was an essential tool employed by contractors to OPS and API as a means of combining a variety of input data sources in order to identify and map drinking water and ecological USAs. The contractors used a leading GIS software package, Arc/Info, and wrote programs in Arc Marco Language (AML) to automate the use of Arc/Info.

How the technical review was conducted

Barney Krucoff attended a workshop for technical reviewers hosted by OPS and API. The workshop included presentations on the intent of the USA program, the definition of USA, and sessions with the contractors (Research Planning Inc. and McCulley Frick & Gilman Inc.) during which the ecological and drinking water USA models were explained. Baker was also provided with extensive documentation on the USA initiative and with copies of the GIS data and AML code developed for the three-state pilot study. Mr. Krucoff and Wing Chong reviewed the project documentation, particularly the report titled A Model for Defining Unusually Sensitive Areas Under the Accountable Pipeline Safety and Partnership Act of 1996. Mr. Krucoff and Mr. Chong then reviewed each of the AMLs provided. This was done by reading the AMLs and at the same time opening the data files created at each major step of the AMLs using ArcView. ArcView allowed the results of the AMLs to be understood in detail without rerunning each of the models in Arc/Info, which would have made it difficult to meet the scheduled deadline.

Review of AMLs

As part of the review, a check was conducted for some of the basic characteristics of good AML programming. The AMLs from both Research Planning, Inc. and McCulley Frick & Gilman Inc. exhibited the following: 1. They well written and professionally commented, which made it possible to read and follow each program; 2. the Arc/Info computing environment was properly established; 3. although some file names and program perimeters were "hard coded," by and large excellent use was made of variables; 4. the AMLs included error handling; 5. variables were returned to the memory pool once the programs had finished execution.

The AML algorithms were checked against the narrative description of the USA models and in each case the narrative description of the computer code is in proper agreement. The algorithms include particularly imaginative use of the Arc/Info region feature class. Finally, the spatial accuracy of the USA data was checked by plotting the USA data on USGS quad sheets and performing a common-sense visual inspection. In each case the USA data passed the visual inspection. Samples of the check plots are attached to this report.

Recommended Nationwide USA Identification Workflow

The AMLs written by OPS and API's contractors represent a high standard of professional practice. However, the work process used to complete the pilot study will prove very difficult to sustain as the effort is expanded to the entire country.

State-to-state variations in available input data pose a major challenge. During the pilot study, contractors handled this problem by creating separate AMLs for each USA model for each state. This resulted in 6 primary AMLs (2 models * 3 states) and 4 or more subsidiary AMLs per state. The differences within the AMLs from state to state are significant. Each AML is a custom product. This approach, could prove very difficult to sustain as the program is extended to the remainder of the country. For example, if 100 AMLs (2 models * 50 states) were created and then the definition of USAs was changed (even slightly), OPS would be left with a software maintenance challenge. Each of the 100 AMLs would have to be opened, reviewed, understood and modified in order to implement the change.

The central recommendation is to not create custom GIS code (AMLs) for each state and rather to develop two robust GIS applications (drinking water and ecological). The applications would be used for all of the remaining states. This will further require that standard formats for input data be identified in a data dictionary and that all incoming data be converted to the standard formats during a data preparation phase. This process will greatly reduce the amount of GIS code that must be written and maintained.

Data prep can still be automated, but any custom code would not be built into the drinking water and ecological GIS applications. The contractors performing data preparation will have the flexibility to decide on the most efficient means of performing the data prep and this should reduce the overall life cycle costs for the project.

It is also recommended that OPS consider developing a preliminary risk assessment data set be created and distributed with the USA data sets. The data set would be designed to aid pipeline operators in interpreting the USA data by showing them what areas are immediately uphill or upstream of USAs. A digital elevation model (DEM) (commonly available GIS data from which slope can be derived) and hydrographic data could be used to map areas immediately uphill or upstream of USAs. The map would highlight areas from which product from a pipeline accident could potentially reach a USA. This is important because Pipeline operators will be instructed to consider USAs if their pipeline crosses a USA (easy to interpret) or if damage from a pipeline accident could reach a USA (more difficult to interpret). The preliminary risk assessment map would tell pipeline operators right away if a nearby USA is within potential reach of liquids from their pipeline.

Impact of recommendations of USA Project Schedule

The impact of these recommendations on the project schedule will be a key consideration for OPS. The development of a USA data dictionary and robust GIS applications that can be used to create USAs for the remaining states will require time and effort. To save time, OPS could continue developing USA data using the pilot study workflow while developing the new GIS applications. If OPS chooses to develop one application at a time, then the ecological model would be the first choice for automation because the incoming ecological data is more uniform from state to state.

Work Flow Diagrams

Diagram 1 below summarizes the work process used to create the pilot study USAs. *Diagram 2* shows a four-stage workflow that will prove more sustainable as the project expands nationally and would provide additional functionality for pipeline operators.

Diagram 1. Pilot Study USA Workflow

1. Gather Data
Data is collected.

→

2. Prepare Data
Significant data preparation occurs. However, to the extent possible, the data collected is left in its native formats.

Data prep is not uniformly as well documented as other aspects of the pilot study. Sometime data prep occurs before the AML is run sometimes it is included in the AML.

→

3. Write custom AMLs and produce USAs
Custom GIS applications are developed for each state, reusing code to the extent possible. The custom AMLs handle variations in the incoming data by state.

→

USAs

Diagram 2. Recommended Four-Stage USA Workflow

1. Gather Data
Data gathering would occur as it did during the pilot study.

Key recommendations

- *A data-gathering manual should be created.*
- *Minimum standards for input data quality should be established.*
- *The requirement to collect ecological data 5 miles into neighboring states should be dropped.* *

→

2. Prepare Data
Data is converted into a standard USA input data formats. (This is a major change from the pilot study, where data was to the extent possible left in its native format.)

Key recommendations

- *Create a USA data dictionary. The data dictionary should describe acceptable input and output data formats in detail.*
- *Convert all incoming USA data into formats established by the data dictionary.*
- *Continue to use contractors with subject matter expertise. Require them to use professional judgment and to document their work.*

→

3. Analyze Data Using a Standard GIS Application
The same application(s) / GIS code is used for every state. (This is a major change from the pilot study, where custom applications where developed for each state.)

Key recommendations

- *Develop new GIS applications that analyze data in the standard data dictionary formats*
- *Include a graphical user interface that walks the user through the process of loading the standardized data.*
- *In the ecological model, during multi-species analysis, include USA candidates from neighboring states as that data becomes available.* *

→

USAs

↓

4. Create Preliminary Risk Assessment Data
Use a digital elevation model (DEM) and hydrographic data to map areas uphill and upstream of USAs where product from a pipeline accident could reach the USA.

Key recommendation

Pipeline operators are instructed to consider USAs if their pipeline crosses a USA (easy to interpret) or if damage from a pipeline accident could reach a USA (more difficult to interpret). The preliminary risk assessment map would help pipeline operators interpret the USA data by highlighting areas immediately uphill or upstream of USAs.

Data Gathering Recommendations in Detail

1. Data gathering should occur as it did in the pilot. The contractors did a good job of finding and identifying data sources.

   - When the USGS National Hydrographic Dataset (NHD) is completed, it should be used instead of the USGS DLGs since upstream/downstream information will be available, thereby increasing model accuracy. NHD data is already available for the majority of states, and a test of its usefulness to the USA project could be performed.

2. A very brief data-gathering manual could easily be created from the pilot study documentation and would be helpful to those carrying on this work nationwide.

3. OPS should also consider standardizing minimum scale/accuracy requirements for incoming data. Currently the ecological model prequalifies species point data based in part on its spatial accuracy. However, this is not done for drinking water intakes or wells, perhaps because spatial accuracy data is not available.

4. USA contractors should be provided with the same state boundary files used by the National Pipeline Mapping System (NPMS). For the pilot, they used ESRI files. The boundary files used by NPMS are created by the Bureau of Transportation Statistics and will be made available on the NPMS Web site (www.npms.rspa.dot.gov).

5. Data from neighboring states is required by the ecological model during multi-species analysis. However, collecting and processing neighboring state data within a five-mile buffer requires as much work as collecting data for an entire state. Given that some states have numerous bordering states (Tennessee has eight), this process becomes unworkable. The recommendation is to charge contractors with collecting data and processing USAs one state at a time. As neighboring states become available, multi-species analysis would be rerun for areas within five miles of the border, perhaps resulting in the identification of additional USAs.

6. Contacting every public water system in order to determine whether or not an alternative water supply is available is a major challenge. Perhaps there are other federal agencies that would also like to know more about the public water systems, and would share the burden of this effort if their questions were included in the census.

Data Preparation Recommendations in Detail

The biggest challenge faced by the USA project is that input data varies from state to state. During the pilot study this was handled by creating custom AMLs. Under the recommended workflow, the customization would occur during the data preparation phase. Thus, under the recommended workflow the consultants will spend more time preparing data and less time customizing AMLs. The overall workload will be about the same, but the GIS code will be more stable.

1. OPS should continue to employ contractors with subject matter expertise (ecology and geology). Because incoming data varies from state to state and unique situations will continue to arise, firms with environmental as well as GIS expertise should continue to do this work. The firms should be allowed to use professional judgement in how they handle variations in incoming data. In exercising that judgement the firms should be instructed to make conservative decisions and err on the side of identifying too many USAs rather then not designating one that should have been identified. They should also be required to document the data preparation work for each state.

2. In moving to the recommended workflow, a set of specifications must be developed for the USA data dictionary and for the USA GIS applications. A first step will be creating a list of the data that the applications will require (e.g., aquifer outcrops) and a list of the data that, if available, the application must be able to use (e.g., aquifer subcrops).

3. A GIS data dictionary should be developed for the entire USA project. The data dictionary should define standard file naming and data structures for all input and output data. For example, take the National Heritage Program data. Texas and Louisiana use very similar NHP data formats and California's data is somewhat different. Assume that the Texas and Louisiana format was adopted as the USA standard. The USA National Heritage Program data would look like the sample below. All Arc/Info coverages and tables used in the USA project would have a similar entry in the USA data dictionary.

Sample USA Data Dictionary Entry

*Naming Convention: "**_nhp" (where ** = two letter state code)*
Format: Arc/Info
Feature types: Point or Polygon
Description: Standard USA input format for National Heritage Program data.
Table Descriptions

Item Name	Width	Output	Type	N. Dec	Indexed	Valid Codes	Comment
**_NHP-ID	4	5	B	-	N		Point ID
ELTYPE_CODE	1	1	C	-	N	A=(A)nimal, P=(P)lant	Species Type
ELCODE	10	10	C	-	Y		Element Code
SNAME	80	80	C	-	Y		Species name
GRANK	10	10	C	-	N	G1..G5	Global Rank
PRECISION	1	1	C	-	N	M=(M)inute S=(S)econd G=(G)eneral (U)nkown	Mapping Accuracy
SoOnAndSoForth..							

4. Incoming data will be converted into the input formats described in the USA data dictionary. It will not be necessary to delete fields from the incoming data, but it will be necessary to make sure that all the fields in the data dictionary are present and that they are coded according to standard. This will require interpretation of the incoming data, which is why subject matter experts should be involved.

   For example, the item *PRECISION*, which measures mapping accuracy, is included in the sample input data dictionary above. In the pilot study, Texas and Louisiana had a *PRECISION* field matching the data dictionary description, but California used a different field, *ACC_CLASS*, which had a 1 to 10 coding system. Under the recommended workflow, the contractor would add the *PRECISION* field to a working copy of the California data and assign a code--M, S, G, or U--based on his professional judgment. Under the pilot study workflow, the contract modified the California ecological AML to handle *ACC_CLASS*. The workload is essentially the same, but fewer AMLs are created.

Data Analysis Detailed Recommendations

Rather than creating separate AMLs for every state, OPS should develop two generic USA GIS applications (drinking water and ecological). The goal would be to develop robust GIS applications capable of running the models for every state without modification. It is assumed all incoming data would be made to comply with the USA data dictionary. The GIS applications, if developed, should have the following features:

1. The application will include a graphical user interface (GUI). The GUI will help guide users so that, to the maximum extent possible, the USA models are applied in the same way from state to state. The GUI will also allow layman to better understand the model.

2. At the start of the model run, the application will ask the user a series of questions and the answers to those questions should be recorded. The questions would be designed to provide the computer with enough information to run the model for the state. Some questions would be simple: for example, “What state/area is being studied?” Others questions would be more complex. For example: “Does the state have defined well-head protection areas or will default values be employed?”

3. Importantly, the application will have a series of browse buttons that will aid the user in loading the data to run the model. For example, a user running the drinking water application would be asked to:

   - click “browse” to load your point coverage for public water systems
   - click “browse” to load your polygon coverage for aquifer outcrops
   - if available, click “browse” to load your region coverage for aquifer subcrops
   - etc.

4. File names should not be hard coded into the application. Instead, variables should be used to the maximum extent.

Conclusion

The USA pilot study has proved to be a very useful endeavor. The hard work of defining what USAs are and creating computer algorithms capable of identifying them has been accomplished. Even if the definition of USAs changes somewhat based on the comments from other technical reviewers, a sturdy foundation has been created.

This review focused on the use of GIS. The central criticism is that the pilot study work perhaps did not put enough emphasis on the problems likely to be encountered when the program is expanded nationwide. The recommendations proposed here are designed to help OPS maintain the level of quality established by the pilot study as the program grows.

In summary the major recommendations are:

- Develop two standard GIS applications (ecological and drinking water) and use them for remaining states. Phase in this approach as necessary to stay on schedule.

- Handle discrepancies in incoming data in a more formal data preparation phase. Develop a USA data dictionary to guide this work.

- Continue to use contractors with both GIS and subject matter expertise.

- Consider helping pipeline operators interpret the USA data by also providing a preliminary risk assessment map.

- Prepare a brief data gathering manual.

- Use the same state boundary files as the NPMS.

- Perform multi species analysis across state boundaries only as data in neighboring states becomes available.

- Look for partner agencies to aid in contacting public water systems.

- Consider setting minimum quality standards for incoming data.