Using Technology to Change Work: Technical Results from the APA Prototype

CTG.APA-014

Center for Technology in Government
University at Albany / SUNY

Table of Contents

Executive Summary

The Adirondack Park Agency (APA) administers land use regulations for approximately 6,000,000 acres of public and private property in northeastern New York State. In particular, the Agency reviews proposed projects for compliance with various environmental, zoning, and statutory requirements. To do this, staff assemble and review numerous paper documents, maps and related materials. They also analyze precedent documents drawn from the Agency's historical archive of paper and microfilm.

On average, the initial review of projects takes 30 or more days to complete. Collecting the appropriate data to support a review consumes much more time than the analysis of the project itself. In most cases, the analysis and determination take only a few days. Gathering materials, rendering geographically oriented data into a consistent scale, and moving files between staff in the agency take much more time.

The Center for Technology in Government, located at the State University of New York at Albany, worked with the Agency to develop a prototype system to combine document records and geographic data into a unified workstation. Using this prototype, the Agency was able to identify those areas where technology would improve their responsiveness or reduce operations costs. The implementation also illustrated for the agency the financial and staff cost needed to develop software and convert data from various sources.

This report presents the findings of the technical staff responsible for developing the prototype system. A combined geographic and document database was not difficult to prototype. Much of the geographic data was already available in electronic form, or was easily converted from other formats. Documents about parcels were also available in non-electronic formats. The project included activities designed to allow the agency to evaluate different processes that would involve either selective or comprehensive scanning.

Data conversion was much more expensive, and required much more time, than originally planned. Gaps and inconsistencies in data indexes made integration difficult. Geographic references, the key relationships between datasets, were often missing or inconsistently coded. Microfilm archives of documents contained many more images than were needed on-line, and did not distinguish between historically important and unimportant information. Other microfilm materials were on media which required specialized equipment unavailable to most vendors.

The prototype implemented only part of the original APA proposal, concentrating on data integration and user access, and removing workflow functions. Working with a visually-oriented, user-friendly workstation provided great insight into the requirements, priorities and functions of a future system. Hands-on experience with data conversion gave an appreciation of the need to improve current record-keeping

and data standards. The users were able to fully participate in the modeling of costs and benefits based on experience, rather than conjecture.

The lessons learned about data organization, technology implementation, and the relationships between technology, process, and people are applicable well outside the scope of this project. Our recommendations on technical requirements for the APA's future GIS and document system include:

- Integrate changes in the agency's work processes with changes in technology.
- Provide user access to technology, and remove organizational and technical barriers.
- Develop small-scale applications rather than attempting comprehensive solutions.
- Implement stronger data and infrastructure standards to ensure reliable and consistent systems.
- Continue to cooperate with outside agencies in the development and exploitation of new data sources.
- Carefully evaluate the amount of data to be converted for the system.

1. Overview and Background

This document presents the findings of the CTG technical staff responsible for developing the APA prototype. It summarizes the technical approaches used during the project, some of the obstacles we faced, and our conclusions.

From the outset, the primary objective of the APA project was to demonstrate a rapid document and map retrieval system for all Agency records relating to real property. The information was to be keyed to property and owner information as well as to conventional Agency transaction codes. This system would allow Agency staff to respond immediately to public inquiries that presently require extensive and time-consuming data collection and would eventually allow remote access to essential Agency documents through inexpensive digital copies and/or online access.

In the spring of 1994, the Adirondack Park Agency proposed that the project prototype consist of a fully automated Jurisdictional Inquiry (JIF) process, with electronic transfer of work products through the organization. [1] JIF processing was targeted because it is similar to other processing at the agency. A Jurisdictional Inquiry comes from a landowner or other interested party. To respond to the inquiry, staff members at the agency collects information from maps, paper records and microfiche documents. This information is sent on paper to the individuals who contribute to a determination of whether the agency has jurisdiction over the project. Once completed, the finding is delivered, and the reference documents are filed for future reference. With an automated system, decisions related to real property for which repeated future reference is important. This would include for example, permits, letters of non-jurisdiction, all determinations that a development project or subdivision is "vested," formal determinations of the existence of a wetland, and agreements settling violations of the statutes administered by the Agency are instantly retrievable on-screen.

The agency expected that “[d]esign and evaluation of the prototype will...provide a budget and specification for Agency-wide implementation of such systems.” [2] Corporate partners were invited to participate in this activity, and Computer Sciences Corporation was selected to serve as principal systems architect for development of the prototype. Hewlett-Packard Corporation provided development and evaluation platforms for this activity, and NYNEX Corporation donated telecommunications resources.Environmental Systems Research Institute, contributed ArcView2, a front-end for its Arc/Info software. Excalibur Technologies Corporation provided use of its EFS document management product.

Development of the prototype comprised several activities:

---

[1] Memorandum from John Banta, Adirondack Park Agency to Peter Avery, CTG, March 24, 1994, p. 1
[2] Op. cit., p. 2

1. Creation of an Integrated Geographic Information and Document System. This involved creation of a on-line access system that provided APA staff and constituents with access to geographic data, geographic analysis tools, and paper documents such as deeds and project files. Given that no commercial off-the-shelf system offered this combination of functionality, the prototype entailed building a custom integration of a Geographic Information System (GIS), and a document management system.

2. Populating the System with Agency Data. The agency's existing electronic databases of geographic and record-oriented data needed to be modified to support the new on-line functions. In addition, paper- and microfilm-based documents were converted into electronic form and entered into the prototype database.

Because of the volume and variety of data needed to support JIF processing (see CTG.APA-009), it was not possible to create a database containing all necessary information for the entire Adirondack Park. Given that the agency wanted the prototype system to be data rich, it was decided to incorporate all necessary data from one area of the park, rather than providing a more extensive coverage of partial data for a larger geographic area. Essex County was chosen as the prototype target, with two local regions (North Elba and Lake Placid) having the most complete data available.

The remainder of this paper describes the processes and lessons learned from the prototype activities. Section 2 presents our approach toward designing and implementing the prototype database. The issues related to creating our prototype GIS/document management system are discussed in Section 3. The paper closes with a short list of recommendations.

2. Constructing an Effective Database

Integration of geographic and document data requires an understanding of the logical and physical relationships between data types and sources. Within a single technical platform, such links are readily available. A primary function of a GIS is integrating disparate data through programming or visual display. Similarly, document management systems allow the creation of logical folders, which join together individual pages into groups for manipulation. When joining different physical data types, conflicting organizing principles make inter-relationships more difficult to exploit.

While database modeling may clarify theoretical relationships, in practice these relationships may still remain cloudy. At the APA, we found that the lack of standardization in data coding further limited the effectiveness of the system.

In the following section, the logical data model for the APA prototype is presented. By examining the model, we identify those areas where we expected data integration to be feasible. In section 2.2, we discuss the base of GIS data that was used to begin the prototype. Following this section, we turn to the efforts of the technical staff to develop a useful database from the independent data sets maintained by the agency.

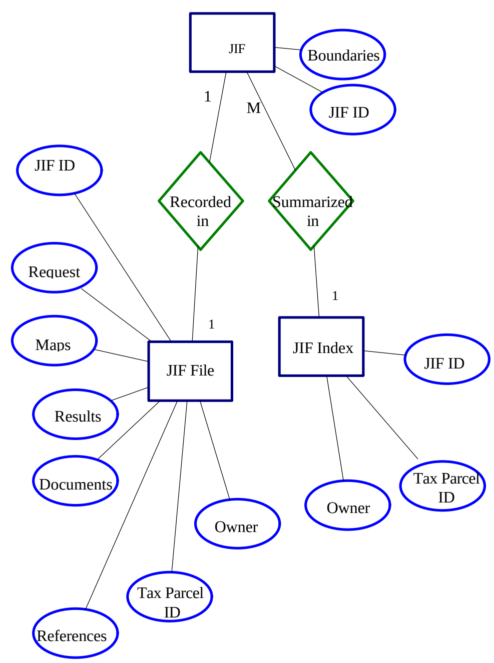
2.1 APA Logical Data Model

The APA logical data model is divided into sections for Jurisdictional Inquiries, Projects, and Maps.

Jurisdictional Inquiries

Jurisdictional Inquiries (JIF) are formal requests to determine if a property falls under the APA's purview. The JIF covers a specific land area, defined in the data model as a boundary (Figure 2-1). The request is accepted in writing from a developer or landowner, recorded in a JIF File, and assigned a sequential number.

Figure 2-1
APA Data Model - Jurisdictional Inquiries



The JIF File contains information on the ownership and location of the property under examination. It contains paper records of the request, and supporting documentation: maps, documents, plans, and other materials collected or created during the agency's analysis. In the model, Owner represents other identifying names which may occur in the data, such as corporate owners and applicants (persons who are applying for a JIF but do not yet own the property). The location of the parcel is presented as Tax Parcel ID, which uses a state-wide standard tax locator. Street address or general location (e.g., "North of Route 86") is also part of the file. A JIF may reference other JIFs.

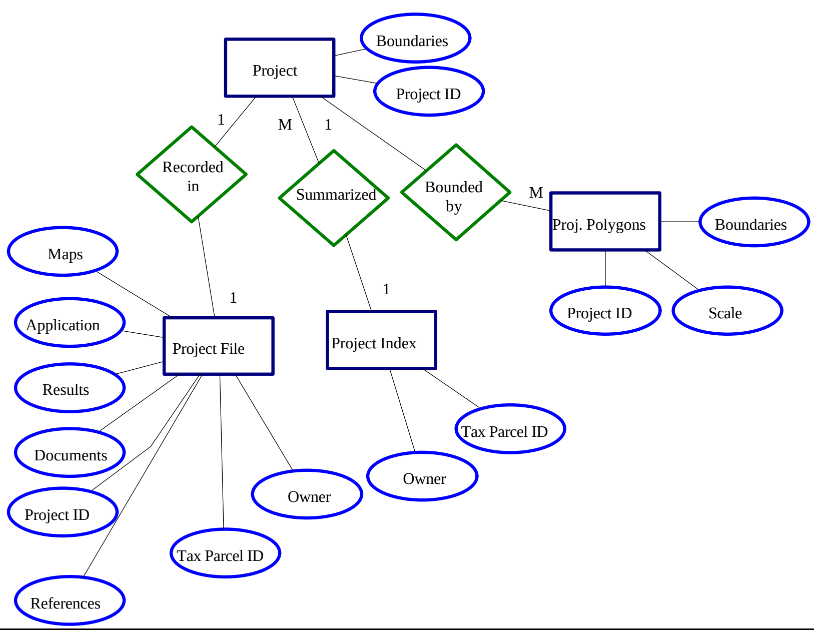
A second record of the request is created in the JIF Index. This is an electronic list of the key elements of the inquiry: ownership, location, and other variables. The JIF Index is maintained independently of the case folders through a PC-based program. One entry in the JIF index may reference several JIFs.

Once the JIF determination is made, the results are sent to the person who initiated the action, and a copy is filed for future reference. The JIF index is also updated. In most cases, the JIF results in a finding of non-jurisdiction; that is, the agency asserts that the property does not fall into any of the classifications under its control.

Projects

If a property is found to be within the jurisdiction of the agency, a project is defined. Figure 2-2 presents a model of this portion of the agency's data.

Figure 2-2
APA Data Model - Projects



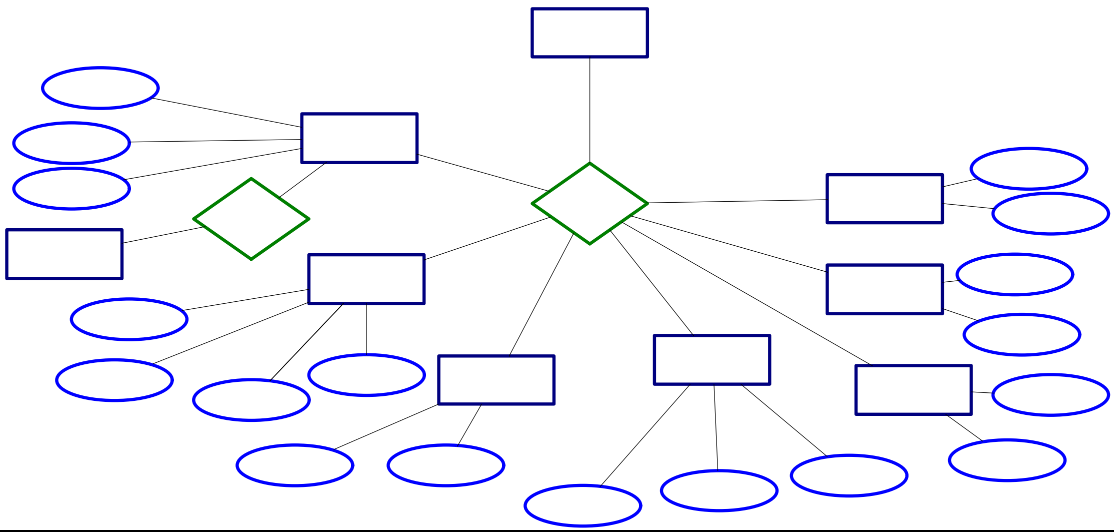
The project data model is very similar to the JIF model. Again, the paper documents for the project are recorded in a project file, which is maintained independently of the JIF materials. The project data is also summarized in a project index, whose characteristics are similar to the JIF index.

The third relationship in the model defines the link between a logical project and its composite project polygons. The polygon entity comprises the graphical representation of the boundaries of the project stored as an electronic map. In its electronic form, the polygons have dimensions, as well as a project identifier. As with all spatial data, the polygons are implicitly defined with a scale reference, indicating the limits of precision in coding.

Maps

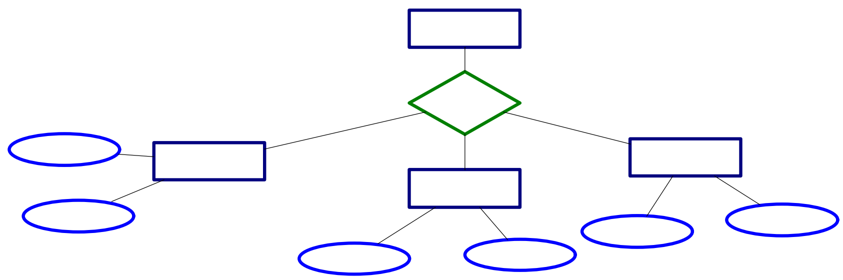
The third data type in the model is the body of map data available to the agency for its analyses. The agency uses many types of maps to describe the park and its contents. Figure 2-3 presents some of these. In addition to the project polygons, there are geographic representations for about 30 data types.

Figure 2-3
APA Data Model - Map Data



Parcel centroids depict the approximate center of tax parcels; centroids are identified by their location, and are coded with their scale, ownership, and tax parcel ID. Land use maps define the boundaries of areas with restricted development. Municipal boundaries include both counties and municipalities. Wetlands and aquifers are examples of a large group of environmental coverages. Road locations are also available. A second set of map data includes planimetric maps, tax maps, and topographic maps (Figure 2-4).

Figure 2-4
APA Data Model - Map Data (Continued)



We distinguish these two sets of data for a pragmatic reason: While logically the second group has similar characteristics to the first, their electronic representation is different. The first group is stored electronically as vector data: these elements may be manipulated and analyzed through the geographic information system. The second group is raster data: it is stored as a collection of grid cells, and may be used as an underlay or overlay for vector data, or simply viewed as an image. It cannot otherwise be manipulated.

Implications of the model

The logical data model presents the complexity of relationships that exist among APA data. For example, given a single project ID, we have determined both the project file and the associated M polygons. Within the geographic data, most of the relationships are many-to-many (M:N).

The model also lists a number of attributes shared by several entities, these give us the ability to establish new relationships between data. For example, owner data is stored in project files, JIF files, their associated indices and parcel centroids. We can use this data to establish a common index for all agency actions. Tax Parcel ID is a refinement of Tax Map ID; these two elements can also be used to link centroids, tax maps, projects, and JIFs.

The logical modeling of data is a useful exercise to understand basic conceptual relationships. It does not show which relationships are exploitable under a particular technology set. This is important when attempting to combine software products and platforms in unusual ways. When we combine the logical model with knowledge of the data manipulation abilities of a GIS, we can greatly improve our integration in ways not well explained by the model. For example, we can determine if one element of a vector data set overlaps another. This means that, given a project's boundaries, it is possible to identify distances to aquifers, wetlands, or other environmental sites.

With digitization of document files, these relationships become even more valuable. Identification of a project allows the location and display of the documents associated with the file.

Using a data model

The data model presented above was not put to paper until the project was completed. Components of it were developed during the project: we understood the limitations of the linkages between geographic and file data early in the analysis, but formal description of the model was documentary, rather than part of the analysis process.

Why did we not build one earlier? Since we relied heavily upon an existing GIS implementation to build the prototype, we used the data models it was built upon. Extensions to the model were constrained by the capabilities and limitations of the products we used. A more extensive project, integrating workflow and other new data structures would have required a data model to facilitate an efficient design.

Though the data model, underlying structures, and implications were clear to the software developers on the project, the absence of a paper model was felt when others tried to understand the functions of the system. In retrospect, an articulated data model, even one that was limited in its depth, would have provided important background for others.

2.2 Integration of Existing Data

Approximately 80% of the total cost associated with a Geographical Information System is related in some way to data collection (Thapa and Bossler, 1992). [3] Because it is both costly and time-consuming to locate and acquire data, convert it to digital form, and integrate it into a GIS, the prototype included as much of APA's existing digital data as possible. Coverages present in the existing GIS were integrated into the prototype without conversion, and several database files were imported directly from their native structure. Many types of image data (TIFF, Erdas LAN, and others) were also used in their original formats. In this manner, the prototype became functional far more quickly than if new data had to be acquired from outside sources. Table 2-1 lists some of the data initially included in the prototype; Appendix A, Table 1 presents a comprehensive list of data and sources.

[3] Thapa, Khagendra, and John Bossler, "Accuracy of Spatial Data Used in Geographic Information Systems," Photogrammetric Engineering and Remote Sensing 58:6 (June 1992), pp. 835-841.

Table 2-1
Sample Data Layers - Existing Data

Data Type	Scope	Scale
Parcel Centroids	Essex County	1:24000
Real Property Parcel Data	Essex County	1:24000
APA Project Sites	Park-wide	1:24000
Land Classification	Park-wide	1:24000
Land Classification (Large Scale)	Park-wide	1 pixel = 1 acre
Biological Sites	Park-wide	1:24000
Road Network	Park-wide	1:250000

Spatial Data

One common problem with the use of spatial data, determining its level of accuracy, was minimized since many of the coverages used in the prototype were acquired from state agencies such as the Department of Environmental Conservation and the Department of Transportation. These coverages are widely used throughout New York, and a high level of accuracy is assured since these agencies may be held liable for any error or inaccuracy in their data. Coverages created by APA were assumed to have an adequate level of accuracy for agency needs.

Temporal inaccuracy or inconsistency may become a problem when working with spatial data. An especially difficult situation became apparent when tax parcel data from NY State Office of Real Property Services, Project and JIF tables from APA, and tax maps from localities were compared. Each of these datasets was updated at a different point in time; thus ownership information was not consistent from one to another. Exact parcel location and extent were likely to be inconsistent as well. The most recently updated of these three datasets would show new lots or sublots, and consequently new owners, that the others would not.

Variable scale across coverages is an unavoidable problem when working with spatial data. In this project, scale accuracy problems were evident when combining displays of some of the ecological coverages with parcel tax maps. Most coverages were created at large scale (1:24000), and were displayed at an unacceptably large scale only when used as overlays on tax maps, where scale ranged from 1:1200 to 1:9600.

We attempted to ensure spatial accuracy by registration of all spatial data to a consistent base map. In the prototype, data was frequently referenced at the tax parcel level. Existing tax maps might have been a good choice for the base map, but acquisition and conversion of data at such a large scale was far beyond the scope of the prototype. Instead, the base maps used were 1:24000 planimetric maps created by NY

State Department of Transportation, which were converted to digital format by the NY State Department of Environmental Conservation. Although no "official" base map for New York State exists, this base map may be considered a *de facto* standard, since it is widely used throughout the state by both public and private sectors, and is of sufficiently large scale to serve as a base map for a wide variety of applications.

Tabular Data

In addition to the spatial data added to the prototype, several of APA's existing database tables were added, to be used as indexes to both spatial data and document images. Four tables were included: one each for APA Project Sites, Jurisdictional Inquiries, Enforcement Sites, and Pre-Existing Subdivisions. Prior to their addition to the prototype, these database tables resided on four separate computers, and were maintained by different departments. Because of inconsistencies, numerous problems were encountered in the use of these tables as indexes, ultimately limiting the functionality of the prototype.

Initial prototype design specified three ways to locate a tax parcel: by tax ID, landowner name, or geographic location. Once a parcel was selected, a map showing the location of the selected property could be displayed, and Project Site or Jurisdictional Inquiry (JIF) documents pertaining to the parcel could be viewed. In order to perform these tasks successfully, the unique tax ID associated with a parcel was to be used as a link between the table containing parcel data and the project site and JIF tables. We assumed that a parcel could be uniquely identified by tax ID in all tables where it was present.

The only table containing detailed information about individual tax parcels was provided by NY State Office of Real Property Services as a component of one of the existing coverages. Each record could be identified by tax parcel ID, and the table could be queried by all three required methods. For reasons detailed below, this table could not be directly linked to any of the APA's database tables.

Each of APA's database files was organized by file ID number, meaning that this was the *only* value in a table that could be relied upon to uniquely identify a project site or JIF. Identification by tax ID was not possible, since a value for tax ID was missing in over half the records in each table. Records which contained a tax ID did so in a variety of formats, and any or all of these formats might be encountered in a single table. Project sites might be composed of one or many tax parcels, so selection of a single project ID did not guarantee that a single parcel would also be chosen. Lists of tax IDs for each project site were stored in a separate database file which was not available for the prototype.

Querying project site and JIF tables by landowner name was also impractical, since the distinction between landowner and permit applicant was unclear, and values for both landowner name and applicant were often missing. Query by geographic location was possible only for project sites, and this was done by linking the project site table to spatial data for each site through the project ID field. In any case, none of these queries produced the unique tax ID needed to link either table to tax parcel data. We addressed this problem by creating an indirect spatial link between tax parcel and project site, overlaying parcel points on project sites, and visually determining which project site(s) applied to the chosen parcel.

Lessons Learned

From the experience of gathering together and integrating existing data, several lessons were learned:

Use of existing spatial or tabular data is a simple, quick, and cost-effective method of populating a new GIS with data. This is particularly desirable when prototyping, since it enables a prototype system to become functional in far less time than when data must be acquired from outside sources.

Spatial datasets (coverages) and tabular data are, in most cases, easy to convert to the desired format and integrate into a GIS. This is true despite the fact that most GIS's store their spatial data in a proprietary format. Most GIS's have extensive conversion facilities for other major GIS formats, and database software is normally capable of converting to and from a wide range of tabular data formats.

The temporal aspects of data are an important consideration, especially with tax parcel data. Updates concerning changes in tax parcel information, such as subdivision of lots or changes in ownership, must be made consistently to all pertinent datasets to avoid the creation of an inconsistent and misleading database that cannot be relied upon for either spatial or temporal accuracy. However, when using data provided by others, control over the timing of updates is rarely possible. Thus, a degree of inconsistency within the system must be tolerated. The presence of this inconsistency and its effect on the conclusions drawn from the data must be made clear to all users in order to minimize potential confusion.

Metadata is a valuable addition to spatial data that makes sharing data with others easier. Metadata is simply data about other data. It provides a concise summary of the information a user needs to know about a spatial dataset in order to determine its suitability for a given purpose. Such information might include the dataset's creation date, its creator, the primary data source, the original scale, and intended use of the dataset. At present, the addition of metadata to spatial data is considered optional, but its inclusion is becoming much more common, and should be strongly encouraged. Knowledge of a coverage's metadata enables users of spatial data to make informed decisions as to whether or not dataset is the right one for their particular purposes. In addition, as agency needs and personnel change, the original structure and assumptions about the data may be lost if it is not well-documented.

Capturing metadata also permits the agency the information to provide support for unforeseen data needs.

Scale variability is unavoidable when combining a wide variety of coverages of differing types. All spatial data is created at a specific scale, and is most accurate at that scale. Although it is acceptable to use a coverage at a smaller scale (features reduced in size) than the one at which it was created because spatial accuracy does not suffer, it is *not* acceptable to use a coverage at a scale far larger (features enlarged in size) than its original scale. A general rule of thumb is that the largest scale at which a coverage may be displayed while maintaining spatial accuracy is twice the original scale: for example, a planimetric map whose original scale is 1:24000 should not be relied upon for a high degree of accuracy when displayed at a scale larger than 1:12000. It is important to note that decisions made based on overlays of coverages displayed in violation of this rule must be tempered by an awareness of the potential for inaccuracy.

The base map used in a GIS will ideally be of a scale at least as large as that of the most commonly used coverages. Larger scale base maps can provide a high degree of detail and accuracy, but the drawbacks of increased cost of data acquisition and conversion, greater data complexity, and increased time required for update and maintenance of databases must also be considered.

Conversion and integration of digital data into a GIS may be a simple task, but the utility of this data once integrated is by no means guaranteed. Before attempting to integrate tabular data, particularly if it is to be used as an index, the data must be evaluated carefully to ensure that it is capable of supporting the required functionality. Data modeling is a valuable tool for examining the existing relationships between data elements, and showing where others might be created. Although a data model cannot show all aspects of the relationships between data elements, it can show the explicit relationships required between different types of data, and suggest the means of establishing these relationships if they are not already present.

Data that was previously adequate for a given purpose may be all but useless when used in a GIS. The implicit relationships that are readily understood by people who work with data on a daily basis must be made explicit when the same data is used in a GIS, since it cannot "see" the relationships. For example, in the APA prototype, the database files would be more useful as indexes if they were normalized and re-indexed with tax map ID used as the key field for each table. The necessary link between parcel points and project sites or JIF sites could then be established. Missing values could also be added to all tables to enable queries based on landowner name or other fields.

2.3 Adding New Data

We now turn to the particular problems faced in the conversion and development of new map and document data for the project. An important part of GIS implementation is the conversion of data to digital form, and integration of digital data into a single repository. These steps are equally important for document imaging systems. In each case, data from a wide range of sources, often in many scales, sizes, and formats, must be converted to a format compatible with the chosen software and hardware. Table 2-2 lists some of the new data that was converted and added to the prototype. Appendix A, Table 2 presents a comprehensive list of data and sources.

Table 2-2
Sample Data Layers - New Data

Data Type	Scope	Scale/Density
Wetlands	Town of Keene	1:24000
Tax Maps	Town of N. Elba	1:24000
APA Project Sites	Park-wide	1:1200 - 1:9600
JIF Documents	Essex County	300 dpi

Analog data such as maps, drawings, blueprints, paper documents, microfilm, and microfiche may be converted to digital form in a variety of ways. The fastest and lowest cost conversion method is usually scanning. It is well-suited for the conversion of paper documents, but may be less desirable for converting maps, since the resulting files are stored as raster data, as opposed to the vector data most commonly used in a GIS.

Digital data such as spatial datasets, coverages, or database files are usually more flexible than analog sources. Here, the challenges arise when the data is stored in a proprietary format. Converting these datasets to alternative formats may range from easy to impossible, depending upon the data's original format.

New Maps

Several tax maps for the town of North Elba were added to the prototype datasets. These had to be scanned at the Essex County Real Property office by APA staff, since the original maps could not be removed. The maps were originally large scale Mylar maps, at scales ranging from 1:1200 (1 inch = 100 feet) to 1:9600 (1 inch = 800 feet). These were scanned at 400 dpi using a Vidar TruScan large document scanner connected to a 486 class PC, and integrated into the prototype as Group IV TIFF files.

Topographic maps were also scanned by APA and added to the prototype. Topographic maps are similar to the planimetric maps used as base maps for this project, but include contour lines to show elevation or topography for a quad. These were done in-house from map sheets already in APA's possession. Scanning was done with the same hardware described above, and was also stored as Group IV TIFF data.

Paper Documents

Two different approaches were taken to converting paper documents for inclusion in the prototype. The first technique was a comprehensive scan of pages from a set of project files selected by the agency. Later in the project, another set of project documents was scanned, using selected file contents.

Comprehensive scanning. The first attempt at data conversion was a simulation of an in-house manual conversion effort, using PC-based technology. In May, 1994, a clerk spent two weeks scanning a set of project folders from Essex County parcels. The first 25 projects from 1973 were used; these folders ranged from a few to several hundred pages. These documents included deeds, hand-drawn maps, staff notes, and any other materials stored in the folders. No attempt was made to eliminate redundant or unnecessary documents. The scanning was performed on a 486 class PC, using a desktop scanner (HP ScanJet IIcx) and Windows software.

During the test period, several hundred documents were scanned. On average, about 150 pages per day were scanned, compressed, and stored in this desktop experiment, for a production rate of about 4 minutes per image (assuming a six hour work day). Computer processing took about 1 minute per page at 300 dpi, while the mechanics of paper handling (removing staples, paper placement, re-assembling folders) took much of the remaining time.

For a small set of documents, an attempt was made to implement optical character recognition. Processing time increased markedly, with only fair results. These results may have been caused by the choice of technology, or from the variable quality of the source documents. Since optical recognition was not a primary goal of the prototype, this path was not pursued further.

Selective Scanning. Later in the project, a second attempt was made at in-house scanning. This time, agency staff reviewed the contents of 50 projects near Lake Placid and North Elba, which evolved as the focus of the prototype dataset. Projects were selected from each available year (1973-1994), in order to provide a chronological view of the project process. From each project, items were selected which were representative of the most commonly encountered types of documents.

Following this review, 250 pages were scanned, again at 300 dpi. When compressed in Group IV TIFF format, most documents required 50 - 60 KB per page. These pages were stored with a fixed naming convention, which included the project folder number. This structure was used in turn to load the documents into Excalibur's EFS document management software for retrieval.

Various agency staff reviewed the results of the scanned documents under EFS, and felt that the selected resolution was adequate for their purposes. While reading documents on-line was helpful, it seemed that users would likely choose to print the scanned images.

Microfiche Documents

Much of the staff time spent on data conversion concerned the transformation of previously archived data to electronic form. As part of an earlier document archiving effort, the Agency had reduced part of its historical database to analog media. JIF files from 1979 through 1985 were recorded on 16mm microfiche. In addition, some JIF files and project data (ranging from standard page sizes to D- and E-size architectural drawings) had been transferred to 35mm film.

Since the manipulation of this data required specialized equipment, both types of data were outsourced to service bureaus for processing. By the end of the project, the 16mm data had been digitized, and Essex County data made available in the prototype. The 35mm-based data had not been processed, for reasons discussed below.

16mm data conversion. At the outset of the project, the agency estimated that about 17,500 pages were available on fiche. Though all the documents in this dataset were still available in paper form, there was no catalog of images. Several vendors were contacted for bids, and were given sample fiche for experimentation. For the final processing, the agency located uncut fiche rolls which were used as the final data source. The vendors were asked to provide an index to documents and folders to our specifications. The index was constructed from identifiers found on each fiche image in the form YY-9999, where YY was the year of the transaction, followed by a transaction number assigned chronologically by APA.

Somewhat surprisingly, the final image count was over 35,000. Many of the film records included both sides of a document; when a page was date stamped on the reverse, for example, the stamp was duly recorded on film. Most of the original data was scanned from 8 1/2” x 11” typewritten documents, and converted legibly.

The indexing of images was more difficult. Spot checks of the initial set of scanned images against the indexes prepared by the vendor showed several errors and inconsistencies. The vendor made several attempts before we felt secure in the indexing effort. There were also several problems with the formatting of the tapes used to move data from the vendor’s computers to the project workstations. In the end, however, no comprehensive test was performed on the data.

These images were also stored as Group IV TIFFs, with 50-70 KB per page, similar to the manually scanned project files. Once loaded, the images were found to be of good quality and legible on the workstation screens. Four months elapsed between the selection of a vendor and the completion of document processing.

35mm data conversion. A secondary project goal for the APA was the conversion of its base of 35mm microfilm records to digital form. Unlike the JIF data, these images were either no longer available at the agency or were inaccessible. Thus, conversion of this data was important both to the project and to the Agency as a whole.

Unfortunately, this effort was not completed during the project. Most of the vendors who responded to the 16mm bid request could not process 35mm microfilm. Those which responded had difficulty reaching the quality standard desired by the agency.

Why was this dataset a problem? There were great deviations in the types of documents recorded on the 35mm film, ranging from standard paper formats to E-size architectural drawings. The recording density (and the associated reproducibility) varied as well to accommodate these different formats. The quality of the original documents was also suspect. Again, there was no reliable index to the documents to determine what data was stored.

Vendor pricing varied greatly as well. Larger format documents required more customized processing by the vendor, with a necessary increase in price. The uncertainty in content resulted in great variations in pricing estimates.

By the end of the project, two vendors had been located which were able to process the 35mm data to the standards needed by the agency. This information was given to the APA for subsequent review and procurement. Eight months elapsed from the time the first vendors were contacted until this point had been reached.

Unresolved Conversion Issues

Obtaining and Maintaining Tax Map Data. Most of the data sources used in the prototype are maintained by State or Federal sources. Tax map information, however, is compiled by county governments. It is probably the most dynamic geographic dataset in the prototype, as changes in property ownership are reflected there (as well as in the state-level Equalization and Assessment data). At this time, only a few of the counties in the park have converted tax map data to digital form. Most have not yet begun this process.

There is an opportunity for the agency and these local governments to work cooperatively to develop and maintain these data over time. The APA has begun working with Essex County to gather its first set of databases. Similar efforts may be facilitated by the work of the New York State GIS Clearinghouse, currently underway at CTG.

Variation in technical specifications. From discussions with scanning vendors, we learned that there is a great variability in document compression and resolution standards. It was fortunate in that the particular technologies used in the prototype (ArcView2 and EFS) were able to use a common variant of the Group IV TIFF format. Without a common standard, additional conversion would have been necessary prior to loading digitized data.

Conversion planning. Recommendations for several important implementation issues were based on the results of the prototype conversion efforts. Cost tradeoffs, and priorities for data conversion are presented in Andersen, et al., “Interim Report on the APA Data Population Model,” (CTG.APA-009).

The time and effort required to specify the document data conversions were well beyond original expectations. Due to the limited scope of the prototype, no attempt was made to perform rigorous quality control on the results. Thus, we recommend a very conservative plan for estimating a timetable for conversion of the remaining documents.

Lessons Learned

From the experience of converting and integrating map and document data, several lessons were learned:

Scanning of tax maps is an inexpensive first step to getting cadastral (tax parcel) data into a GIS. However, for maximum utility, such maps should be vectorized, or converted from raster to vector format. Scanned, or raster images, incorporated into a GIS may be used as display images only; they cannot be used for spatial analysis techniques such as overlay analysis, area calculations, or boundary checking. Raster data is stored in a grid cell pattern, where map features are approximated by filling in the grid cells that most closely resemble a feature. Vector data is stored as a series of X, Y coordinates, and the location of every point along a feature may be determined. Because every point can be known, continuous lines or polygons may be constructed whose length, area, perimeter, and exact location may be calculated.

Service bureaus that perform document conversion provide highly specialized services; the services offered, quality of results, and price charged are likely to vary widely. For these reasons, it may be necessary to choose vendors based upon their ability to produce high-quality output from a single type of data. Sending sample data to several vendors is highly recommended; in this way image quality, job pricing, file storage requirements, and timeliness of delivery may be evaluated. Several vendors, each with a specific area of expertise, may be needed to complete the conversion process.

File formats for document images are an important consideration. Group IV TIFF is a commonly used file format, but there are many subformats which may be vendor or hardware-specific, and these subformats are not necessarily compatible with one another or with the software used to display document images. Choice of vendor may be strongly influenced by the output file formats they are able to provide and their ability to be read by image display software.

Tight quality control is needed at every step of the conversion process. This will ensure the best possible image quality, as well as correct indexing of all document pages. Indexing problems may be very difficult to detect and correct; providing the vendor with an accurate index of microfiche or microfilm contents is a necessary preventive step.

Storage requirements for document or map images are likely to be substantial. Despite the fact that the Group IV TIFF format includes file compression, the large number of documents a typical system would include makes it necessary to estimate the number of images to be stored, and the size of each file in order to acquire the appropriate type and quantity of storage hardware. Some average figures taken from the prototype are listed in Table 2-3.

Table 2-3
Data Types and Document Sizes

Data Type	Approximate Document Size
Document images	50-60KB per page
Tax maps	125KB per map sheet
Planimetric maps	300KB per quad sheet
Topographic maps	3MB per quad sheet

3. The Prototype System

During the project, two prototype systems were developed. This section of the report discusses issues raised during the development of these systems. A detailed timeline describing the prototype development is contained in Appendix B.

3.1 Issue 1: Scope of the Prototype System

Computer Sciences Corporation and the agency originally envisioned a prototype that contained a complete set of functionality for JIF processing. The original design [Computer Sciences Corp., 6/1/95] included customized user interfaces for different staff members depending on their job functions, with workflow capabilities that would route JIF applications through the appropriate processing steps. Scanning and printing modules were included so that all members of the agency who currently handle the paper files could understand how the technology would impact their jobs. This system was to be built by integrating four technologies: a geographic information system, a document management system, a workflow system, and a database system to serve as data repository and communication link between system components.

When it became clear that resource constraints would preclude constructing a prototype to the original specifications, many changes were made in the design.

Workflow capabilities such as on-line forms and routing of folders between staff members was eliminated from the prototype. Adding new documents by scanning them in was not possible in the prototype, and all documents were batch-loaded. A single user interface combining functionality common to all users was provided. Because of the smaller scale of the prototype, data management tools provided by the GIS and document management systems were used in lieu of a specialized database system. The technical details of the prototype system are contained in Appendix C.

In October 1995, an initial prototype was developed for the Interchange '94 conference, and shown to selected members of the staff. This prototype did not integrate the two technologies, but had separate GIS and document management systems containing a portion of agency data.

In hindsight, this simplified system was exactly the right thing to provide to the agency. It allowed staff to see the basic capabilities of the technology at a relatively low cost, and to imagine how the technology could be used to do their current jobs better. More importantly, because agency staff were exposed to the same functionality, they could envision how technology could enable work redesign by combining tasks and re-organizing workflow. If we had constructed specialized user interfaces for individual staff, this would have served to automate the current division of responsibilities in processing JIFs. Instead, this sample of the basic capabilities allowed the agency to think more broadly about how the technology would change their operations.

3.2 Issue 2. The End-User Platform

A second major issue faced in the prototype design was the selection of an appropriate end-user platform. While the envisioned technologies were UNIX-based, many systems could work with a Microsoft Windows or X Windows front-end, allowing the user platform to be a PC.

Both UNIX and Windows had advantages. UNIX front-ends are more commonly available, require less configuration and customization (of both hardware and software), and are typically more stable and full-functioned. A Windows-based front-end, however, would demonstrate to the agency the possibilities for integrating the proposed system into other work areas of the agency: word processing, electronic mail, and access to other agency databases. A Windows-based environment would also be substantially less expensive than one using UNIX workstations.

In the end, we decided to implement a UNIX front-end because of availability of hardware and software platforms, and of support personnel.

3.3 Issue 3: Integrating GIS and Document Management Systems

The prototype system is described in Appendix C. Based on Hewlett-Packard's HP-UX version of the UNIX operating system, the system used three software components. ARC/INFO 7.0.1, a robust fully-functional GIS, was used to create the data used in the prototype. It was not used by end-users, but was used in an off-line mode to organize geographic data to be used in the system. The primary user interface was provided by integrating ArcView2 for viewing of geographic data, and Excalibur EFS 3.5 for document management capabilities.

The original plan was to create a separate main user interface in C which would drive the two packages to provide the necessary functionality, with a DBMS to add data storage support. Part way through the project, however, ArcView2 was released. It provided a flexible user interface with appropriate UNIX hooks to allow it to serve as the primary user interface, and consequently we chose it to serve as the main control module in the application.

The integration of the two systems was very loose. An ideal linkage between modules would have been at the subroutine level, by calling appropriate Application Programming Interface routines in Excalibur. Unfortunately, although Excalibur had a user interface and data model appropriate for the project, the product did not have an API library for system development and integration. As a consequence, communication between the two applications was done using a command-line interface and shell scripts. Essentially, the ArcView2 application launched an Excalibur session; once started, communication between the two components of the application was non-existent.

This led to a prototype application with several undesirable features. Of the most importance was the inability to create and store new work products. For example, a user could locate a parcel, and view the documents associated with it, but could not store these data together into a new folder. The prototype was implemented as a 'read-only' system, and could not accept new relationships among existing data.

A related problem was the absence of relationships between groups of documents. Each logical document file was implemented as an independent entity within the document management software. A file name was constructed using the agency's coding structures for its ongoing work; there were no contextual links (e.g., Owner's Name) between documents. The prototype did not take advantage of many of the possible links between documents across folders, and relied on the GIS software to display overlaps and relationships.

The absence of interprocess communications between the GIS and the document software caused several implementation problems. The most important was the inability to change the focus of the document software once it had been started. For example, once a particular parcel's documents were displayed, the user could change their position in the GIS. While a new parcel was displayed in the map window, the document window still showed a page from the original parcel. The GIS could initiate a new session of the document software, but could not re-use the old session to look at new documents. This lead to a buildup of open sessions which quickly slowed the prototype's performance.

In spite of these limitations, the integration of technologies that could display maps and deeds in a single integrated system was critical in educating the APA staff in how the technology could revolutionize their operations.

3.4 Lessons Learned

A quickly-developed prototype is better than one with more features that takes longer to construct. A critical step in the development of this prototype was showing an initial version to the agency within months of the original design meetings. Although this prototype did not integrate GIS and document management as they envisioned, merely introducing the agency staff to a copy of the technology loaded with their own data was a powerful catalyst to their imagination. Subsequent project activities, including the final prototype design and the modeling and benefit analyses, were much more grounded in reality because staff had seen the technology in action.

When evaluating a product, the unavailability of a "perfect" technology can be compensated for by other activities. In this project, we depended on the Center's Corporate Partners for technical support in developing the prototype. An ideal technology would have worked on both PC and UNIX platforms, would have a tight integration of the GIS and document management facilities via an API or other form of inter-process communication, and would have integrated workflow capabilities. Due primarily to resource constraints, the prototype delivered to the agency did none of these, yet the project was extremely successful in meeting its goals. The agency has a much better understanding of the technology's potential, and can design the system it will implement with the confidence of having experienced first-hand how the technology works.

The lessons learned are not always the ones you set out to learn at the beginning. While this is something we observe in all the projects the Center has executed thus far, this was especially true in this project. In the process of working with Computer Sciences Corporation and CTG staff in defining the workflow to be automated, agency staff members had an opportunity to discuss what they actually do in performing their jobs. As a result of this activity, the agency has a better understanding of its work flow and how it can be changed (with or without the technology) to improve service. Many of the results of these re-engineering sessions were precipitated by discussions about what the prototype system should do, which led to greater understanding of what the Agency could achieve through simple process redesign.

A prototype is a valuable tool to focus discussion. Seeing a live demonstration of a technology, with agency data loaded into the system, served as powerful catalyst for discussion of the important issues about the Agency's business problem: What is the problem we're trying to solve? How do we currently attack the problem? What do we need to be able to solve the problem?

While these questions don't require a prototype to answer, it is necessary to ask and answer them to develop a relevant system. In one sense, the prototype software is disposable once these questions have been asked and answered. Nevertheless, having a prototype in hand shows how information, processes and technology work together. As a result, Agency staff are better able to create and evaluate solutions that involve technology.

4. Recommendations

Technical innovation should be evolutionary, rather than revolutionary. GIS and imaging technologies are changing rapidly. We believe that the agency would be best served by developing its data and applications incrementally, rather than all at once. This allows the agency to adjust to changes in standards and requirements over time, and to avoid expensive technology which is quickly obsolete.

Combine process improvement with technology development. One of the early steps in the project was the consideration of a change in workflow, facilitated by technology. While the timetable and resources for the prototype did not permit this to be completed, it should continue to be part of the design of new service delivery mechanisms. More time should be spent looking at how processes are performed, and what data is needed at each step.

Code data consistently. One of the problems encountered in preparing data for the prototype was the inconsistent coding of information within the agency. In its report on the APA's records management program, the State Archives and Records Administration recommended the documentation of a standard set of data field names, contents, and use for the electronic records at the agency [4]. We concur. We believe that the experience of the prototype has shown how useful such consistency can be to establish relationships between data elements, and how the lack of coding consistency limited the effectiveness of the system.

Note though, that SARA's recommendation on linking information by tax map number is somewhat difficult to implement. First, the data model shows that tax map numbers, and even parcel IDs do not uniquely identify a project: a project may have several parcels, and a parcel may be in several different projects. Second, tax map information can change over time, in ways out of the control of the APA. Third, there are hundreds

---

[4] New York State Archives and Records Administration, Analysis and Recommendations Concerning the Records Management Program of the Adirondack Park Agency, March, 1993, Summary Recommendations, pp. 1-3.

of records for which this data is not present. We would refine the recommendation to state that tax map information should be captured for all transactions, and that database technology can be used to combine this data with other elements to create unique identifiers.

Establish a reliable technical infrastructure. As the Agency's use of technology increases, so does the cost of recovery from system failure. The agency should establish the procedures for backup and restoration of electronic data. Power supply backups should be installed on hardware and data servers. Staff should be trained on the importance of data and site security. Off-site storage of backups should be established.

Continue prototyping applications. Once software development began, the project had some success developing GIS applications incrementally. This process allowed users to react to developments quickly, and provide valuable feedback. We encourage the agency to continue this process.

Avoid isolating data and resources. While technology can be introduced gradually, we believe that the ultimate target should be universal access to data across the agency. GIS and document applications should be designed to support a wide group of users, rather than a smaller group of specialists. Particular attention should be paid to including persons with limited exposure to technology.

Continue to develop cooperative scanning. During the prototype, the Agency established a mutually beneficial relationship with Essex County to scan tax map information. This was a very useful effort, and demonstrated the value of cooperation between organizations in the park.

Pursue selective scanning of image data. During the development of image data for the prototype, entire files were scanned. There was no differentiation between hand-drawn maps and legal determinations. It was observed that only some parts of files (such as permits) needed to be available in scanned form, while others were less useful. If this is remains true, we recommend that a mechanism be devised to delineate the documents of particular importance before sending items out for archiving. This will facilitate subsequent retrieval and recording.

Consider alternatives to up-front conversion for little-used data. The methods and resources spent on data conversion should reflect the use of the data. For example, we encourage the consideration of a "vectorize as needed" approach for data, such as tax maps, that are retrieved infrequently. Other data, such as project polygons, are likely to be used more often. Rather than vectorize all data as it is scanned, we recommend that vectorization take place as needed, based on the type of data involved.

Finish converting microfilm data. One of the objectives of the agency during the prototype was to convert otherwise inaccessible data from its 35mm microfilm. This was a much harder task than anyone anticipated. Much progress was made in this area, including locating and evaluating several different vendors. Circumstances prevented the project team from completing this work. We encourage the agency to finish the evaluation and have the data converted to a useful format.

Retain microfilm reels when creating fiche documents. SARA recommends the use of microfilm and microfiche for archival storage of paper files. Note that when we sent filmed images out for scanning, there was a substantial cost difference between scanning rolled film (which required little physical manipulation) and jacketed fiche. We recommend that if fiche are created, roll films be stored to facilitate later document scanning.

Integrate word processing and other technologies. Adding word processing to the GIS system would provide particular benefit to the APA's clients. In particular, the ability to generate official findings and letters directly from a workstation may greatly improve the turnaround time for customer transactions. It also provides the ability to capture this information directly into electronic files for later use.

Appendix A

Data Summary for APA Project

The original data included in the APA prototype consisted of a variety of datasets already possessed by APA in digital form. These data included numerous ARC/INFO coverages containing attribute information about the Adirondack Park, scanned images of planimetric maps covering the entire Park, a digital satellite image of the Park, and database files containing project site, enforcement, and jurisdictional inquiry information for the Agency. These datasets are summarized in Table 1.

Documents pertaining to Jurisdictional Inquiries and APA Project Sites were added to the prototype during the course of the project. JIFs from Essex County for the years 1973 to 1985 were originally stored on 16 mm film (microfiche), and were converted to Group IV TIFF files by Gannon Technology of Reston, VA. A total of 35,692 pages were converted, at a cost of $0.245 each. The converted files were provided on 4 mm DAT tapes; Perl scripts were used to load these into EFS and assign the correct filenames.

The APA Project Site documents were originally paper documents; fifty important project folders for Essex County, selected by APA, were scanned manually and entered into the prototype by APA. Approximately 250 pages of documents were scanned. In addition to these documents, an ARC/INFO coverage, a scanned topographic map, and two scanned tax maps were added to the prototype. A summary of the data added to the prototype is shown in Table 2.

Various indexes were used to link database files, spatial data, and document images together. For example: JIF documents were accessed based on the JIF number contained in a record selected from the Jurisdictional Inquiry index. The value contained in the JIF_number field of the record was used as an argument when EFS was started, which then brought up the appropriate document images. Project documents were accessed in the same manner, using the Proj_number field. Enforcement documents were not included in the prototype; the enforcement index, in this case, was simply a viewable table, as was the pre-existing subdivision table. Each of the above index files was originally a compilation of APA's transactions of a specific type, stored in dBase format.

Several additional indexes were also added to the prototype. These were database files associated with Arc/Info coverages, and were added as needed for use with ArcView2 scripts; except for one table, none of these were seen or accessed directly by the users. The exception was the Attributes of Essex County Parcel Points table. This could be searched to find a specific land parcel or parcels.

Table 1: Original data layers:

Data Type	Format	Data Source	Scope	Scale	Filename(s)
__Geographic Data:__					
Planimetric base maps	Arc/Info polygon coverages	NYS Dept. of Environmental Conservation	Parkwide	1 : 24,000	various coverages[1]
Thematic Mapper image	unclassified raster satellite image	EOSAT (Earth Observation Satellite Company)	Parkwide	1 : 1,000,000	tm1429may.lan
USGS names	Arc/Info point coverage	United States Geological Survey	Parkwide	1 : 24,000	adknames
Real Property Parcel data	Arc/Info point coverage	NYS Bureau of Equalization & Assessment	Essex County	1 : 24,000	esco
Land Classification - Small Scale	classified raster satellite image	Adirondack Park Agency / EOSAT	Parkwide	1 : 24,000	landclas.gis
Land Classification - Large Scale	Arc/Info polygon coverage	Adirondack Park Agency	Parkwide	1 pixel = 1 acre	quad24.landclas
APA Project Sites	Arc/Info point/line/poly coverage	Adirondack Park Agency	Parkwide	1 : 24,000	pkprcl
Significant Biological Sites	Arc/Info polygon/line coverage	Adirondack Park Agency	Parkwide	1 : 24,000	biositeapa
Natural Heritage Sites	Arc/Info point coverage	NYS Dept. of Environmental Conservation	Parkwide	1 : 24,000	nh92
ALSC Pond data	Arc/Info point coverage	Adirondack Lakes Survey Corporation	Parkwide	1 : 24,000	alspoint
Aquifer data	Arc/Info polygon/line coverage	Adirondack Park Agency	Parkwide	1 : 250,000	aquifer, aquifersplit
Ecozones	Arc/Info polygon/line coverage	NYS Dept. of Environmental Conservation	Parkwide	1 : 500,000	pkecoz
Municipal Boundaries	Arc/Info polygon/line coverage	NYS Department of Transportation	Parkwide	1 : 24,000	pkmuni
Hydrology	Arc/Info polygon coverage	NYS Dept. of Environmental Conservation	Parkwide	1 : 24,000	pkhypoly
Essex County Parcel Centroids	Arc/Info point coverage	NYS Bureau of Equalization & Assessment	Essex County	1 : 24,000	esco
Road Network	Arc/Info line coverage	NYS Department of Transportation	Parkwide	1 : 100,000	pksfrd
Shaded Relief Map	Arc/Info GRID data	USGS Digital Elevation Model Data	Parkwide	1 : 250,000	pkshd
__Database Files / Indexes:__					
APA Project Site Index	dBase file	Adirondack Park Agency	Essex County	none	prj_essx.dbf
APA Enforcement Site Index	dBase file	Adirondack Park Agency	Essex County	none	enf_essx.dbf
APA Jurisdictional Inquiry Index	dBase file	Adirondack Park Agency	Essex County	none	jif_essx.dbf

Notes:

1. Coverage names are in the form x99xxcl, and are matched to TIFF files with the corresponding filename xx99xxpl.tif.

Table 2: Data Added During Project:

Data Type	Format	Data Source	Scope	Scale	Filename(s)
Geographic Data:					
Regulatory Wetlands	Arc/Info polygon coverage	Adirondack Park Agency	Town of Keene	1 : 24,000	keenewet
Topographic Base Map	raster scanned image	APA / NYS Dept. of Transportation	Single quad	1 : 24,000	c25swtopo.tif
Tax Maps (2)	raster scanned images (400 dpi)	Adirondack Park Agency	Town of N. Elba	1:1,200 - 1:9,600	154089042001.tif 154089042006.tif
Document Images:					
Jurisdictional Inquiry (JIF) Documents	raster scanned images (300 dpi)	Adirondack Park Agency	Essex County	none	various TIFF files[2]
Selected Project Documents	raster scanned images (300 dpi)	Adirondack Park Agency	Essex County	none	various TIFF files[3]
Additional Indexes:					
APA Pre-existing Subdivision Index	dBase file	Adirondack Park Agency	Essex County	none	pre_essx.dbf
Attributes of Adknames	Arc/Info attribute table	Adirondack Park Agency	Parkwide	none	internal Info table
Attributes of APA Project Sites	Arc/Info attribute table	Adirondack Park Agency	Parkwide	none	internal Info table
Attributes of Essex County Parcel Points	Arc/Info attribute table	Adirondack Park Agency	Essex County	none	internal Info table
Attributes of Land Classification	Arc/Info attribute table	Adirondack Park Agency	Parkwide	none	internal Info table
Attributes of Municipal Areas	Arc/Info attribute table	Adirondack Park Agency	Parkwide	none	internal Info table
Attributes of Quads	Arc/Info attribute table	Adirondack Park Agency	Parkwide	none	internal Info table

Notes:

2, 3. Filenames are in the form yy-9999_99x, where yy is the year, 9999 is the transaction number as assigned by APA, 99 is the sequence number for this transaction, and x is a character indicating document type (Project, JIF, or Enforcement). Some Project filenames have an additional character following the transaction number and before the underscore; this was included to match APA's naming convention.

Appendix B

Prototype Development Timetable

The software timetable for the APA Automatic Permit Retrieval System prototype started with the announcement of the project award in November of 1993. It proceeded though design and development during 1994, and installation for the APA users in Spring 1995 (Figure 1).

Project Definition. CTG and APA began a joint review of the project's concept and scope in November 1993. Starting with the proposal prepared by APA, the two organizations established the framework for software development. Since the APA had been working with the ArcInfo product, this was the assumed GIS platform. Other areas, including hardware and document retrieval software were considered open. In December CTG began inviting assistance from other vendors to join the project, and CSC joined the project in January, 1994.

Prototype and Workflow Design. In February of 1994, the project team participated in a series of facilitated planning meetings. These meetings helped the Agency establish a vision of a re-engineered organization, using changes in both technology and workflow to improve service and efficiency. Concerns were raised at that time about the nature and volumes of data, the capabilities of the software and hardware, the Agency's ability to absorb and use advanced technology in its operations, and project control issues.

In March, CSC conducted sessions at the APA site to review these concerns and establish a software target for the prototype. Over a period of three days, each of the Agency sections involved with the project defined a revised workflow for its area of responsibility. At the same time, the project team began assembling the information required to analyze data conversion requirements. In April, CSC presented its analysis of these sessions in the form of recommendations for a comprehensive prototype, including changes in both technology and workflow.

Prototype Software Specification. A preliminary systems design and workplan were delivered to the APA in June. This design was in the form of a text narration, describing a set of operational procedures for a model 'expert worker' of the future, incorporating both technology and workflow improvements. At the same time, CSC developed a list of software modules required to implement the prototype.

During the summer, the prototype definition was re-scaled to concentrate on GIS and Image Retrieval components. The HP hardware platforms were installed, and Excalibur was selected to support the integration of document images.

Initial Prototype Implementation. The software prototype was created in two phases. The initial phase included the conversion of existing GIS data into ArcView, and structuring the cross-references between datasets. A series of menus and scripts were established to link the GIS and image software. Sample image data was prepared to simulate real datasets.

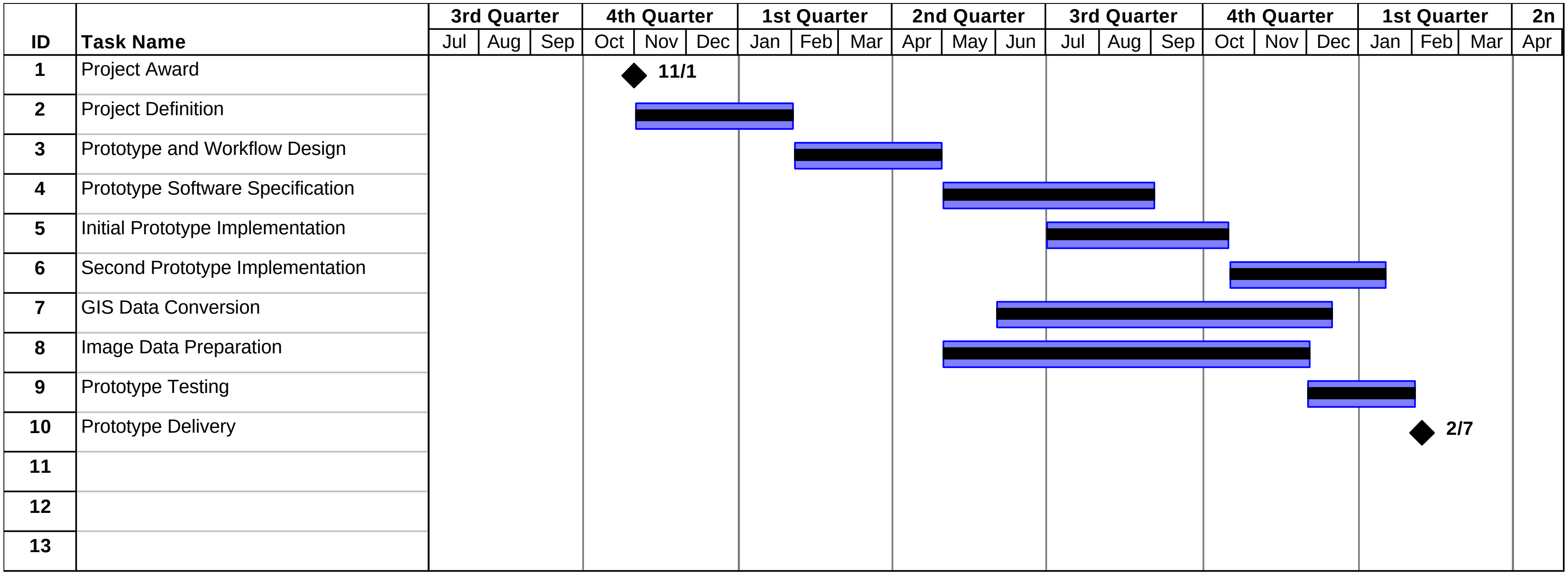
Key users in the APA were shown this prototype of the system in October, and provided comments on its usefulness and areas for improvement. The most important comments for the developers were in the user interface: based on their feedback, the software designers made major changes to the ‘look and feel’ of the system, simplifying access to common functions, and reducing clutter.

Second Prototype Implementation. A second iteration of the prototype was prepared in November and December of 1994. In this version, the software was integrated with the available production datasets of GIS and Images. Many of the recommendations presented by the users from the first prototype were implemented, and a series of software tests were executed to improve system stability.

GIS Data Preparation. Data preparation activities for the project proceeded in parallel with the software development. GIS data conversion began in early summer, as soon as the software and hardware platforms were installed. This conversion, and the subsequent addition of data, continued throughout the prototype. Data volumes required that the team’s efforts concentrate on one section of the park, presenting depth of data, rather than breadth.

Image Data Preparation. Beginning in late Spring 1994, CTG and APA conducted a series of experiments to establish the time required to scan agency documents. A contract for commercial conversion of 16mm microfilm data was signed, and investigations for other conversions continued. Most image data was in place by December, 1995.

Prototype Testing and Delivery. In January, 1995, the team developed a series of training exercises and review materials for the APA’s use in evaluating the project. In February, the prototype was moved from CTG to APA’s offices. There, a series of training sessions were conducted, and the software evaluation phase began.


ID
Task Name
3rd Quarter
4th Quarter
1st Quarter
2nd Quarter
3rd Quarter
4th Quarter
1st Quarter
2n
Jul
Aug
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
1
Project Award
11/1
2
Project Definition
3
Prototype and Workflow Design
4
Prototype Software Specification
5
Initial Prototype Implementation
6
Second Prototype Implementation
7
GIS Data Conversion
8
Image Data Preparation
9
Prototype Testing
10
Prototype Delivery
2/7
11
12
13

Appendix C

Hardware and Software Configuration for the APA Prototype

The hardware consisted of three Hewlett-Packard Model 712/80 workstations, each with the following features: 64 MB RAM, a 1GB internal hard drive, a 1GB external hard drive, and a 20” color monitor, running HP-UX 9.03, and one HP e35 9000 server, configured as follows: 128 MB RAM, 4 GB total storage, a 4 mm DAT tape drive, and one CD-ROM drive, running HP-UX 9.04.

Hardware was configured in two different ways. During the development portion of the project, the server resided at CTG and was connected to the local workstation through a local area network, using 10 Base T. The local workstation was also connected to a Hewlett-Packard ScanJet IIcx scanner. The remote workstation resided at the APA offices in Ray Brook, connected to the server via a 56 kbps leased line provided by NYNEX. Another scanner, a Vidar Truscan 800, was connected to APA’s remote workstation.

During the evaluation portion of the project, the configuration was reversed: the server was now resident at the APA offices in Ray Brook, and the local workstation (now APA’s) was connected to it using thin-net cabling. The CTG workstation, now the remote site, was connected to the server via the 56 kbps leased line. The third workstation, located at Computer Sciences Corporation in Rockville, Maryland, was a stand-alone throughout the entire development and evaluation process.

Both local and remote workstations accessed software located on the server through *telnet* sessions. Software resident on the server included: ARC/INFO rev. 7.0.1, ArcView2, and Excalibur EFS 3.5. Three site licenses were available for each software package. ArcView2 and Excalibur were integrated through the use of UNIX shell scripts. Communication between the two applications was done in asynchronous mode, again through the use of shell scripts and a command-line interface. ARC/INFO was used as a supplementary source of data (coverages) for ArcView2, and was not integrated with the other software packages.

Appendix D

Operational Description of Prototype*

APA Office Support System Prototype

Section 1
Operational Description

This section describes the capabilities to be provided by the prototype system in support of the Public Service Office and Case Worker functions of the APA. The scope of the prototype has been reduced from its earlier concept; the current design focuses on the rapid retrieval of documents and geographic information which was otherwise unavailable or untimely.

1. Screen Overview

When ArcView is started, the user will open an ArcView "project" from the menu bar . An ArcView project is a collection of views (maps), scripts (programs to display data), charts, and tables. (Figure 1). Each of these elements may be accessed by clicking on the column on the left side of the project window. A list of the items available within the type is listed, and may be activated by clicking on the desired item.
When the project is loaded, the Map window is open. It displays an overview of the Park. The Map window has a table of contents on the left border, with the themes available in the database. These themes include tax maps, wetland maps, project sites, and the other geographic data contained in the database. The remainder of the window displays the selected map set.

One or more tables will also be open, presenting the Projects, JIF and Enforcement databases. These databases are presented in a spreadsheet-like form, with each entry as a row in the table.
Above the windows is a set of pull-down menus which provide access to various ArcView functions. Under the menus are two rows of buttons which assist with the most common actions.

2. Public Service Officer / Case Manager

When a PSO or Case Manager is investigating a property or case inquiry, the Office Support System provides access to several types of APA information. Through the workstation, the user can locate a property on a map, and view it in combination with other geographic data. The user may also look for existing related actions, read

---

* This document presents a functional description of the APA prototype as conceived prior to development. The scope of the system was extended based on user feedback.

scanned images from the existing files, and print a record of this information for their use.

2.1 Parcel Identification

The first step in using the Office Support System is identifying the parcel or parcels in question. The system has a record of the Real Property Inventory for Essex County in place, with corresponding geographic references. Locating a parcel of interest may be done in several ways.

> 2.1.1 PARCEL IDENTIFICATION - BY MAP. To query information about parcels in the park, the PSO uses the map to establish its location. When a map is displayed, the user may zoom in or out from the current location, or pan across the map to find the parcel in question.
>
> 2.1.2 PARCEL IDENTIFICATION - BY TAX ID. To locate a property using tax identification information, the PSO selects the Tax Map # Select option from the APA Views menu . This will lead them through a series of steps to identify the town, and map characteristics of the property. The property will be highlighted in yellow on the map window. The scale of the window may be adjusted manually.

Figure 1
Sample ArcView Project Display

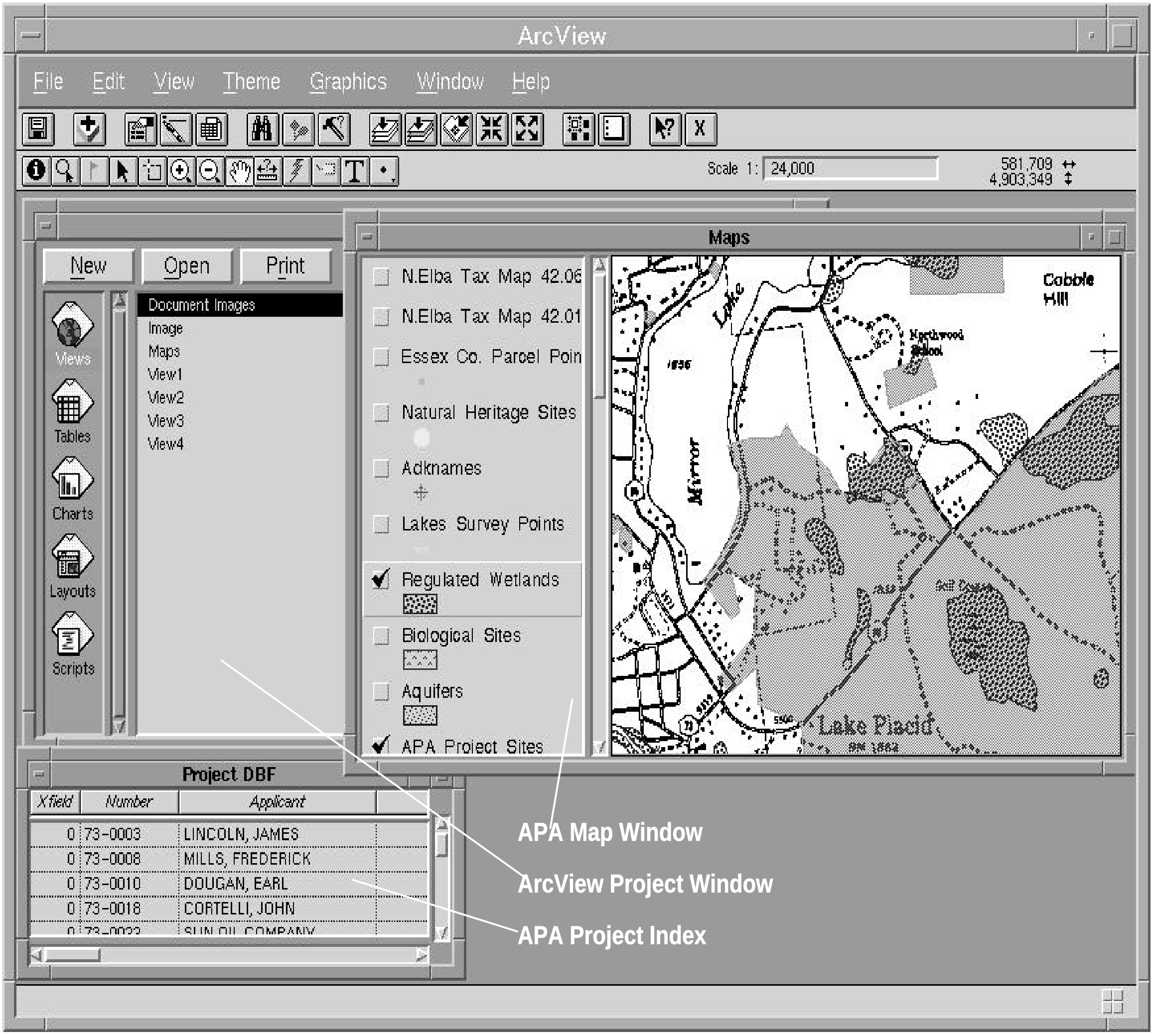
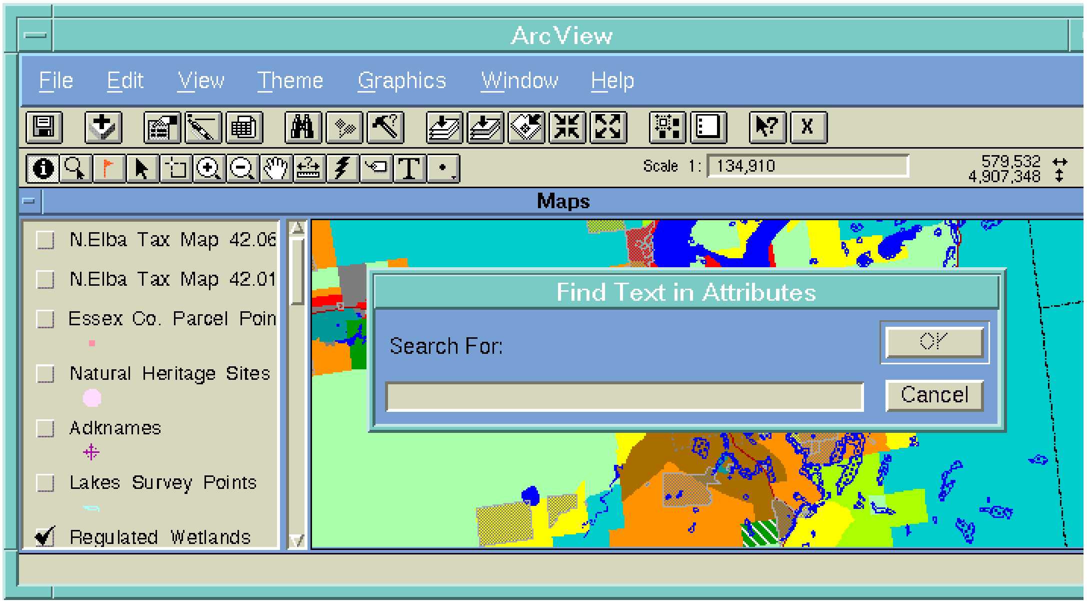


Figure 3

Search of Property Attributes using Find command



2.1.3 PARCEL IDENTIFICATION - BY NAME. A search for a name associated with a parcel may be performed using either the Query Builder or the Find Function. The Find function (Figure 3) of ArcView asks the user for a word or number associated with the property in question. It then finds the first entry in the database which has that word or number anywhere in its record and shows the associated property in the map window. This method allows only one attribute to be specified, and returns only one record.

The Query Builder function of ArcView allows the user to construct a more complicated search. Here, the user may specify each of the values of interest, and the system will return all records which meet the specification.

2.1.4 PARCEL IDENTIFICATION - BY PROXIMITY. To select parcels by distance, the PSO has two options. The first method is to use a 'rubber-band-box'; this allows the user to select parcels from the map window at a variable distance from a given site. Alternatively, the PSO may find all parcels within 100 yards of a selected parcel by pressing the Flag button from the button bar.

2.2 REVIEWING PROJECT, JIF AND ENFORCEMENT DATABASES

The Project, JIF and Enforcement indices as of 9/94 have been loaded into the ArcView project. Through ArcView, the PSO can scroll through the data, or use either the Query Builder or Find commands to locate properties of interest.

2.3 REVIEW OF DOCUMENT IMAGES

Digitized images of documents are available in three ways: based on a map selection, a specific Project Index entry, or a specific JIF entry. Enforcements are not available as part of the prototype.

2.3.1 PROJECT DOCUMENTS BASED ON MAP. From a parcel in the map window, the PSO may choose to view all projects folders associated with the parcel. To do this, they would click on the "X" button. This will start up an Excalibur session each project known for the parcel, up to five parcels.

2.3.2 PROJECT DOCUMENTS BASED ON PROJECT INDEX. To view the contents of a project file, the PSO highlights the relevant project in the Project DBF window, and clicks on the "P" button from the second row of buttons. This will start up an Excalibur session, with the first document in the folder displayed.

2.3.3 JIF DOCUMENTS BASED ON JIF INDEX. To view the contents of a project file, the PSO highlights the relevant project in the Project DBF window, and clicks on the "J" button. This will start up an Excalibur session, with the first document in the folder displayed.

Whenever a folder is opened, the first image of the first document will be displayed. To view documents for the same property in another index, the user will need to select the property and re-query.

The PSO will be able to review or print the documents in this folder. When finished with this folder, the PSO will close the folder manually.

2.4 MAP BUTTONS

The APA database has three predefined views of maps available. These permit rapid access to standard views of data. These maps include:

Group 1: (All at 1:24000) scale	Land Class Planimetric Tax Map Wetlands Parcel Centroids
Group 2: (At 1:2400)	Tax Map Centroid
Group 3: (At 1:2400)	Tax Map Centroids Wetlands Land Use Planimetric

When the appropriate button is pressed, the scale and coverages displayed in the map window will be adjusted to the above settings.

2.5 PRINTING MAP AND PROPERTY DATA

The PSO will be able to print a standard report on a selected property, consisting of the maps from Group 1 and Group 2, above, as well as a list of the property's data from the Attribute table. To do this, the PSO will press the print button from the menu bar.

ArcView also provides the ability to print the currently displayed map window. The PSO may print the currently displayed map window by clicking on the Map window and selecting the "Print" Option from the File Menu. The default options will print the maps; the alternate selection prints the map legend.

3. Restrictions

- The maximum number of document image sessions is five. This means that no more than five folders may be open at a time across all users.
- The initial project will be read-only. Changes to projects will can be made, but they must be saved with a different name.
- Not all geo-referenced data is at the same scale. Caution is required when overlaying coverages with different scales.

4. Reductions in Functionality

The following functions described in the June documents from CSC are not planned to be implemented for December.

- There will be no recording of new data in the prototype. This inc ludes the storage of notes, updates or other annotations
- Workflow functions have been eliminated
- There is no database support for extended queries or folder bundling
- The number of users is limited to the 5 license limit under Excalibur
- The scope of scanned images will be limited to the amount of data available by December 1. Additional data may be added later in the prototype evaluation.

Appendix E

Related CTG Products

1. Andersen, David F., Kristine Kelly, and John Rohrbaugh, "Interim Report on the APA Cost Performance Model," November 3, 1994 (CTG.APA -007)

2. Kelly, Kristine and Steven Hyde, "Preliminary Analysis of Interview Data," March 10,1995 (CTG.APA-008)

3. Giguere, Mark, "Literature Review and Selected Annotated Bibliography-Adirondack Park Agency," June 24, 1994 (CTG.APA-006)

4. Andersen, David F., Steven Hyde, and Kristine Kelly, "Interim Report o n the APA Data Population Model," March 29, 1995 (CTG.APA-009)

6. Kelly, Kristine and Steven Hyde, "Potential Quality Improvements and Extended Benefits of an Office Support System for the APA," May 1995, (CTG.APA-012)

8. Kelly, Kristine and Steven Hyde, "Results from Pre- and Post- Survey of APA Staff," March 1995 (CTG.APA-013)

9. Andersen, David F, Peter Avery, Steven Hyde, Kristine Kelly, Soonhee Kim and John Rohrbaugh, "Evaluating the APA Prototype: Prospects for Providing Cheaper, Faster and Better Services to APA's Customers," October 1995 (CTG.APA-015)

Center for Technology in Government
University at Albany / SUNY
1535 Western Avenue
Albany, NY 12203
Phone: 518-442-3892
Fax: 518-442-3886
info@ctg.albany.edu
www.ctg.albany.edu