

OFFICE OF STATE PLANNING
'96 MAY 31 AM 11 46
DIRECTOR'S OFFICE

FINAL REPORT

**SURVEY OF HAWAI'I GEOGRAPHICALLY BASED DATA
FOR THE OFFICE OF STATE PLANNING
STATE OF HAWAI'I**

Contract #29070

For: James Yamamoto
Office of State Planning

By: Everett A. Wingert
Pacific Mapping Consultants

Abstract:

This report accompanies the printed and digital forms of the Hawai'i State Data Catalog produced in conjunction with Hawai'i data newsletter, *geo-Data Hawai'i*. These products close the Office of State Planning (OSP) Survey of Geographically-based Data conducted by Pacific Mapping Consultants under Contract #29070. The last issue of *geo-Data Hawai'i* has been submitted for printing and will be delivered to OSP for distribution, shortly after this report to complete contract responsibilities.

This document has been prepared to encapsulate the current geographic data situation in Hawai'i and try to look into the future. This look into the future will be done to suggest some of the changes that appear to be waiting in the wings for Hawai'i. Between the present and the future, the report will present an overview of the data collected, the newsletter, and methodology used to complete the work. It is hoped that these perspectives will aid in reforming this work in a different format sometime in the future.

The report will summarize the experience gained in the execution of the contract and make recommendations for steps that need to be taken in the State to accommodate the rapid changes in geographic data that are accompanying Geographic Information System technology implementation. The report will also comment on some of the patterns in the data, and discuss the data format and its problems.

Table of Contents:

Executive Summary	3
Spatial Data in Hawai'i	5
Hawai'i Data Catalog Contract	12
Hawai'i Data Catalog Structure and Contents	20
Methodological Review	27
Epilog	30

Executive Summary:

The Hawai'i Data Catalog Project was an effort to identify all existing and planned spatial and geographic data in and about Hawai'i. The compilation was intended so that all possible data could be brought to bear on the preparation of carrying capacity studies in certain regions of the State. As the project progressed we perceived a strong need to develop an interim method for distributing current or changing information about spatial data to a rapidly growing body of decision makers using geographic data.

To this end an irregularly published newsletter was established to inform and educate interested parties about geographic data. In its abbreviated life, the newsletter changed form and content from what was initially envisioned. As these changes occurred, *geo-Data Hawai'i* began to take on a life of its own, rather than continuing as a subset of the Data Catalog Project. This delayed the attempt to complete the Data Catalog. The delays diminished some of the value of the Catalog in terms of current data since it was not possible to complete it in a short period of time. However, even though it is not as complete and current as would be desired, it is useful as a model for future local metadata procedures.

The Contract:

This contract was authorized as a part of Act 160 in the 1990 State Legislature. In part, the Act stated "A survey and evaluation of existing data bases and other data systems and sources to determine their applicability to SECTION 3 of this Act". Pacific Mapping Consultants was approached to conduct this survey by Mr. James Yamamoto, Administrator of the Project mandated by Act 160.

Originally, the project was to be conducted by a large mail survey and corps of followup interviewers who would collect whatever data were easily available over a brief summer session. The data would then be assembled in a brief report and transmitted to the rest of the Carrying Capacity Analysis.

Data Collection Approach:

Ensuing discussions between Mr. Yamamoto and Pacific Mapping developed a broader based concept of the task and developed an alternative approach that was thought to make the final product generally more useful. In essence, the mail survey approach was discarded in favor of a slower personal interview approach that was expected to generate more profitable data. The same discussions developed ideas of a Catalog compiled in different data base components that could

be cross-linked and searchable in a machine readable form.

The survey was to proceed in three phases, (1) data base development, (2) preparation of a "Contacts" data base -- individuals identified who would be key knowledgeable individuals in helping to locate data, and (3) by extensive interviews, develop a data base of state, local and private agencies involved with the generation and/or maintenance of geographic data and to outline each class of data held by the unit.

Contract Overview:

Very quickly after the field interviews began, the problems and advantages of the intensive interview approach became obvious. Given the general lack of information about the subject of data and mapping in Hawai'i, interviewers very quickly became sources of information, more than gatherers of information. Hours were spent giving information and help to cooperating groups, often with very little information received in exchange. It became obvious that a whole day of interviews might only produce one unique data entry in the data Catalog.

While the data collection continued, much time was diverted to the preparation of a four page illustrated newsletter about geographic data. This publication was designed to help disseminate data of general interest and to assume an educational role by discussing issues of importance to the geographic data community.

The newsletter component of the project began to dominate more and more of the time available at the expense of prompt completion of the data catalog. In addition, the unexpectedly large size of the data available made it increasingly obvious that a complete catalog of geographic data was impossible. This unsatisfactory balance between newsletter and catalog time was eventually resolved by negotiating the specific sample of state agencies to be interviewed.

Conclusions:

The closing date of the long extended contract was precipitated by the reorganization of the Office of State Planning and the budget cuts experienced by all agencies. Therefore, the closing newsletter now in press and the accompanying materials complete this contract.

Hawai'i Spatial Data:

Interviewing for this project began intensively on the Islands of Kaua'i and Hawai'i with irregular available-time interviews on O'ahu. In the beginning, interviews were concentrated in public sector agencies with a sampling of private firms and non-profit groups. Several patterns of spatial data collection and use quickly became apparent.

Pacific Mapping defined spatial data as any form of graphical or digital data that are associated with some type of coordinate system that would allow mapping. The coordinates might be page coordinates, common in many types of drafted and desktop published maps or one of the formal coordinate systems such latitude and longitude or Universal Transverse Mercator coordinates.

In some cases, data associated with a simple location such as Kekaha, Kaua'i or 2617 Honowai Road, Waimea, O'ahu, will also serve as a coordinate for mapping. All data that were associated with some type of coordinate system were called spatial for the purposes of the survey. In addition to conventional spatial data, we included aerial photographs in our spatial data definition since they are the source for so many types of spatial data.

As expected, in many agencies we encountered data that was not mapped, but could be because they were associated with some form of a location attribute. Health data, pollution data, climatological data, for example are often held in a mappable form and were therefore included in the survey.

Concentration of Data:

In the hundreds of interviews conducted, we cannot remember one group or agency that did not use maps or photographs in some way. Many did not consider themselves users of data or generators of data; but in effect were. After some time of talking about their work, a well-worn annotated map or photograph would be pulled out of a drawer or filing cabinet with an "... oh do you mean this kind of thing that 'Sanchez' does?"

Some small groups held unique data but did not think it of any general interest or use. Surely the interview structure missed a great deal of this type of information because of the abbreviated sample, but some did not pass through the net. However, some of these unique types of data could be easily lost if someone does not recognize its value and preserve it.

Many of these finds do not appear in the data Catalog because we did not want the individual who produces the data to be 'badgered'. In other cases, these unique data are too

hard to classify, but the remainder of this type does appear in the Catalog.

These small treasures include the hand annotated topographic maps of Raul in the Maui DLNR. Raul has annotated the vegetation, in broad patterns and detail as well as other interesting physical features on 1:24,000 topographic maps. He is a good field observer and always knows where he is in the field. Data from almost every hike and field job that he has taken for the last 35 years appear on these quads.

Someone in a small office on Molokai has meticulously annotated an old set of aerial photographs with every cultural or infrastructure facility in their area. Every field pump, stock tank, cattle grate and fence gate seems to be included.

Data of these types exist everywhere but there was not enough time to tease them from their hiding places and establish a pointer to it in the Catalog. There is always a danger of losing this type of buried data. One now-retired employee of a Honolulu came into custody of a 5000 print large-scale vertical stereo aerial photograph collection of O`ahu and many of the neighbor islands taken in 1926-29.

When this individual's generous loaning policy was abused by a University professor, the set was taken out of circulation. We found the set a few years ago and were given permission to have it copied. Before funds could be found to make the copies, the individual retired. Another agency then got involved and the set was designated to be a waste of space and was lost again.

As to be expected, many agencies are concentrators or collectors of data. Especially in the current rush to Geographic Information Systems. Some of the agencies like the Office of State Planning, Civil Defense, Land Utilization branch of the City and County of Honolulu, Department of Land and Natural Resources are great "concentrators" of data while generating new data.

Other agencies tend to generate more new data rather than concentrate other agency's data. The Department of Health, Hawai`i Volcano Observatory, the National Park Service, Soil Conservation Service, and the State Surveyors Office, are sources of new or edited data. Other agencies are known primarily for editing or value added work.

Common Spatial Data:

Editing, correcting, and adding data to existing data sets, while valuable, can lead to the situation where slightly different data exist in different locations. When these variants

exist, it was usually not possible to investigate which form of data an agency held. This situation badly needs to be remedied with better data heritage documentation. Almost every state and county agency held some common data sets which they either used alone or to which they added their own information, mainly for internal use.

Almost every office in the State has sets of 1:24,000 US Geological Survey topographic maps in paper, photographic, or film forms. The Department of Health once went so far as to commission the USGS to print all of O'ahu's quad maps without contours for a special project. Some agencies are still using the (years out of print) 1:125,000 sheets for data collection and recording on the island of Hawai'i. Fewer 1:100,000 or 1:50,000 series are in use, but a lot of the now obsolete 1:62,500s are still in daily use. The USGS orthophotoquads, now discontinued, are also common.

The digital forms of these maps are the basis for the State Geographic Information System, but are far less common in agencies that have not integrated AutoCad or Arc/Info into their operations.

After the ubiquitous USGS Quads, next most common data set are the TMK maps produced by the Real Property Assessment Division of the City and County Finance Department. Either in full scale or half scale forms, sets of these maps are found throughout the state in public and private agencies alike. Heavily annotated with agency data, these maps are major source of reference information for Public Works and Engineering agencies state-wide.

Nature of Spatial Data:

– in-house data:

Those agencies that are data producers tend to generate several distinct types of data. Most groups that generate data, do so for strictly in-house use. Most such agencies do not promote general distribution of these data but often entertain distribution requests on a case by case basis. The reasons for non-distribution policies range from liability fears to privacy issues to the need for special training or background to utilize the data.

These types of data range from the Police Department logging the locations of different types of crimes to Public Works noting potholes to be filled. Many of the "map is the law" statutes are based on annotated maps. The Conservation and Resources Enforcement Division of DLNR holds the original drafted 1:24,000 quads that locate the "legal" boundaries between the different Conservation Zones. Hawaii Vol-

cano Observatory often hand annotate paper quads in helicopters before transferring them to scale-stable 1:10,000 sheets for digitizing. Almost every agency has some class of "annotation in-house" map series for internal use and decision making.

– *proprietary data:*

Some of these in-house sets are not distributed for policy or informal reasons, but some other in-house work is restricted for other reasons. The National Park Service often wants to insure the privacy and isolation of sensitive archeological sites by declaring some of their maps proprietary. The Pesticides Branch of the Department of Agriculture keeps their data about the application rates of pesticides to specific fields private. The Nature Conservancy keeps its data on endangered and threatened species locations private so that they can recoup compilation costs by selling the data to developers and other types of EIS and research users.

– *service data:*

In these litigious times, agencies are increasingly concerned about their liability for data made available to the public. However, several agencies still produce data for public consumption. The largest public agency that produces internally generated maps for public use is the TMK mapping branch of the City and County Finance Department. DLNR agencies produce maps for hikers, hunters, and general recreational use. In the private sector, there are many compilations for public distribution, but rarely do they include new data of planning or management interest.

– *"GIS" data:*

In these times of rapid GIS "building", there is a renewed search for data that may be appropriate for planning and management decisions. When individual data sets are added, they may produce sensitivity to a decision set. A map of lava tubes on the Big Island may be of limited scientific interest, however when combined with other information, such as airport runways and construction sites, the combination may become a value source of information for predicting hazardous situations. Data sets that were earlier considered of marginal interest, take on new importance when they can be selectively mixed with other information.

There is nothing inherently different in data produced for a scientific, engineering, legal, or planning use and for Geographic Information Systems. If the data is produced properly according to standards appropriate for their scale, it can be combined for applications within the accuracy limits of the data.

However, it is assumed that data transformed into a digital form uses a coordinate system that is precisely attached

to allow fitting of data sets at similar scales. It is also assumed that map information has been plotted and classified according to standards that will allow overlaying.

At the present, very little data are created expressly for GIS applications although an increasing amount of data is being collected with direct digital use in mind. The first uses to be generating primarily digital data are in the construction and engineering areas which are dominated by CAD software. However, just because data are available in digital form, do not make them universally appropriate for GIS.

The nature of data prepared for multiple purposes and then converted to a digital form is very important and should be a critical element in any data heritage system. Threshold scales should be associated with each form of data and they should be a part of a data's heritage. Data in graphic form can be easily checked for consistency and match, but the same tests that can be applied to graphic data are not applicable to digital data.

Transformation of Data:

Concerns about appropriate data use must be very carefully addressed by the agency creating data as well as those using it. There are different strategies for transforming data from analog to digital. There are data that should not be digitized except for very carefully defined reasons and there are inherently digital data that must be used only at scales appropriate for its characteristics. Scale and compilation are the defining issues, issues that must be carefully addressed in the planning and execution of a project.

– analog data:

Raw data are inherently analog since every measuring instrument is an estimating device. When data are scaled to a map or associated with a coordinate system and classified, they are still conceptually analog. The issue and limitations of use are related to the steps taken to arrive at their storage media.

It is impossible to relate all the cartographic procedures of data representation here, but every data element that is transferred to a map has an accuracy limit associated with both the raw data and the method of transfer. The accuracy limit is a function of what is acceptable at the final analysis scale. A field worker may sketch the location of a heiau on a topographic map according to where he thinks he is. If his map reading skill and dead reckoning is good to a quarter mile, that quarter mile is the assumed accuracy limit of the data.

If another archeologist sketches the same heiau accord-

ing to her abilities, it is assumed to have equal accuracy. This will produce separate uncertainty circles of a half mile. If the plotted heiaus are combined on the same map, the heiaus could be plotted a mile apart. This type of map is inherently analog and should not be converted to digital form because there is no way of telling whether there are one or two heiaus.

– *analog/digital data:*

A cartographer will normally remove such uncertainty knowing about the nature of the field work that produced the data. This is a type of error is the worst type to guard against and the hardest to correct when digitized.

When digitizing an analog map, the digital file cannot have an accuracy figure that is greater than the sum of the field error and map/scale plotting error associated with the original analog map. However, it can be considerably lower accuracy depending upon the operator and method of digital conversion. It is acceptable to digitize data of known inaccuracy as long as that figure is retained with the data.

– *digital data:*

Raw digital data are the most preferable to use for any application, because unless an error is embedded, the characteristics of the information are known from the outset, unmodified by plotting and classification. Field surveyed data is - the only major class of this type of data in use in Hawai'i. It is easy to estimate the accuracy of this type of data and it should be used as base information if one has the choice. However, field surveyed data are too expensive to collect for almost any application beyond construction and legal uses.

The second main digital collection mode is from Global Positioning Systems surveys. The great advantage of this type of data for all applications is that every point's inherent accuracy is known. This knowledge means that GPS data can be always be applied at the appropriate scale. At the present, GPS data is somewhat limited in Hawai'i but it is finding its way into a number of data sets. At the present DLNR is using GPS data in many applications ranging from archeological applications to trail mapping. The State Department of Health also uses these base digital data for many of their field investigations.

Magnitude of Data:

Beyond the more common data sets, it is very difficult to estimate the actual amount of analog and digital data that exists in Hawai'i. Basic analog data probably still dominates the data scene when aerial photographs and conventional mapping sources, historical and current, are included. The data survey

probably only identified a sample of the information available. It is very hard to count data sets since some are very small and localized, and while useful are difficult to compare to the large sets such as the parcel data for the Honolulu City and County GIS.

At the outset of this project, it was expected that the interviews would quickly start to identify duplicate data sets and focus on a relatively compact body of available data. This expectation was in error. When one takes into account the amount of variants of the sets with individual agency modifications, duplication is only true at a very broad level.

When it was decided to limit this sample, large holes in the data set existed. These gaps could not be addressed without a full inventory of each agency which would take weeks in some cases. It is expected that there are thousands of valuable data sets in the large number of research departments in the University of Hawai'i. By a very rough estimate, including small and large data sets and their variants and derivative sets, it is conservatively expected that over 10,000 spatial data sets exist about Hawai'i when all scales and series are counted. To enumerate and describe all of these sets, would take years. It badly needs to be done, but it is a very large task.

Duplication of Data:

Inherent in the magnitude of data being produced in the state, is the potential for duplication of effort for a variety of reasons. The largest reason is simple ignorance of work going on in different places. In 1992, almost across the street from each other, Hawaiian Telephone and the Hawai'i Electric Company in Hilo were both digitizing a Big Island land base for their own separate but very similar use. It has been reported that currently, one division of DLNR is redigitizing the topographic quadrangles and parcel data that is already in place at the City and County of Honolulu.

The same duplication exists in many other settings. Two university departments are independently converting the USGS DLGs to a NAD83 coordinate system, almost like a race, but each organization uses slightly different coding for the streets and streams.

If for no other reason, the goal of the Hawai'i Data Catalog was sound based on the identification of duplication of data alone. However, even knowing the existence of duplicate sets does not prevent their creation since different groups are working on different hardware and software platforms. And there is the problem of groups wanting their own data. Even

identifying duplicate data does not solve the problem.

Work must proceed on the development of a state-wide clearing house for data that can be shared. An investment in time and facilities that would be repaid many times over, for savings in developmental costs.

Metadata:

In 1992, the federal government mandated that every spatial data set produced by a government agency or project funded by the government, must be accompanied by a standardized form that detailed the data's accuracy, source, collection method, etc, on a many page form. Difficulty with complying with this large task has produced many abbreviated metadata standards and compliance is far from complete.

While this documentation process is incomplete, it is a very necessary task as data are produced and disseminated in many different formats. Some of this work has been done for state-wide data but it is largely absent in local data. The State and City and County GIS projects have made progress in developing and codifying data about their data, but it falls far short of what is necessary for the educated use of data. In addition to the need for cataloging data, this need for knowledge of the nature of spatial data must go hand in hand with its identification. It will never be possible to develop documentation on all of Hawai'i's data sets since they have gone through so many generations, but a start needs to be made.

The variety of data and circumstances described above in Hawai'i provided the setting to begin to catalog and classify data. With the range of data types, functions, accuracies, intended applications, different formats and scales, the Hawai'i Data Catalog project was begun. However, as was quickly discovered, the project was cast far too small to make a serious dent in the problems noted above.

***The Hawai'i Data
Catalog Project:***

The Data Catalog Project procedures began to be formulated in Spring 1991 with a series of meetings to define the scope of the project and develop the characteristics of the deliverable products. As originally focussed, the project was to conduct a brief survey of existing digital data to be used in Office of State Planning Carrying Capacity Studies.

In the early discussions, Pacific Mapping requested a

broader scope to include traditional forms of mapped data and aerial photographs. Instead of using a mail survey with follow-ups, we decided to concentrate on personal interviews that traditionally produce far better data. While this method is slower and can cover much less ground, the quality of the data was judged to be tradeoff that justified the adopting the procedure.

Originally, the final report was to be a listing of the data surveyed, but the early discussions produced an agreement that a digital data base form of the survey would allow machine searching for characteristics and make cross-linking of elements of the data base that would make it valuable beyond its local OSP application.

Distribution of the data base was discussed but no conclusions of a venue or method of distribution was decided at that time. However, the software standard of Borland's Paradox3 was adopted since it was in common use in OSP and was easy to convert to ASCII form for distribution to other platforms and agencies.

Contract Beginnings:

In early discussions, it was decided to conduct the survey in a way that would produce three separate but linked data bases. (1) a "Contact" data base of people developed in many ways to provide a first entrance into field interviews. (2) an "Agency" data base that would contain characteristics of an agency or firm's hardware, software, and data handling capability. It was planned to be able to search for agencies using the Intergraph hardware, for example, and then link that to the forms of data in that format.

(3) The main "Activity" data base would list the actual data and record information about the characteristics of the data. This data base would also contain information about data sources, storage formats and data dates, as well as the projects that generated the data. This data base turned out to be similar to the federal "Metadata" form of data information under development at the same time.

While general definition of the formats of the data bases was still in progress, the Contact data base was begun.

Development of the Contact data base:

The Contact data base was developed in different ways on different islands. The first compilation was made by examining several lists developed by the early interest groups forming to advance Geographic Information Systems in Hawai'i. The interest lists were supplemented by contacting many of the

main players in data collection and management on O`ahu and other islands for suggestions of key people, both private and public that had special knowledge in this area.

At this point State and County phone directories were surveyed to locate key agencies and division heads on all islands. After this survey was completed, GTE Hawaiian Telephone books were surveyed for missing agencies. At this point the yellow pages added the names of private agencies in the areas of photogrammetry, planning, engineering, and public service groups. Work was also done to identify the sources of data on Hawai`i in federal agencies on the mainland through a variety of sources.

When field interviews began later, a closing question of each subject interviewed was a request for additional names that might be useful to add to the Contacts data base and this data base continued to grow and be edited throughout the early phases of the project, until the size the data base reached dimensions that made it impossible to cover in the scope of the project.

Personal contacts in cartography and planning were polled on different islands for developing additions to the Contact data base. In the case of Hawai`i, a contractor was hired to evaluate the data base and add to it. This intermediate development of the Contact data base was only used on the Big Island. While the use of an intermediate produced some interesting leads, it did not substantially increase the size of the Contact list.

When the first phase of the Contacts list was complete a cover letter requesting an interview and a response postcard were sent to all listings. As expected the response to this mailing was incomplete, but a number of good prospects were identified and showed considerable interest in the project. Based on the responses, phone calls were made to set up the first set of interviews.

Early Fieldwork:

An early phase of the fieldwork was begun before the actual intensive interviewing was started. A set of test interviews were conducted with known "friendlies" to refine the nature of the data that would be developed by this type of intensive discussion. This phase also saw the development of supporting materials and handouts to explain the mission of the survey and how their data were to be coded.

While useful, this test interview process should have been done at least partly with "outsiders". Most of the "friend-

lies" were very cooperative and we could push for information when they were uncertain about which tack the interview was to take. Very quickly, we found that the format for the Activity data base was ambitious and many of the contacts simply did not know the information without consulting that actual people generating the data.

The format was maintained in the hope of better luck later in the interview process, but we did end up with many blanks and interviewers being forced to estimate some values and lead interviewees too much on others.

With the test phase complete and the format mostly finalized, intensive fieldwork began in early summer. Because Big Island Contact list was checked and modified by a cooperator, interviewing started there. For much of the summer from one to three days a week was spent on Hawai'i, with the O'ahu days spent coding data and conducting some interviews with large agencies here.

We quickly discovered that coding and interpreting the data took longer than collecting it. We also discovered that no efficient pattern of interviews could be developed without hours on the telephone days in advance trying to coordinate appointments. While the serendipitous drop-in during quiet times still produced very valuable data on occasion, most of the days on the neighbor islands were spent waiting and coding data.

The neighbor island pattern that emerged was to make three appointments a day, allowing for the occasional two-hour interview. The rest of the time was spent developing new contacts and making drop-in contacts on the pretense of setting up an appointment. About half of the time, a drop-in would agree to talk on the spot.

There was little pattern to these early interviews, some potentially high-return interviews produced little data, some were the opposite, some were very focussed and useful in time spent and others spent all their time trying to figure out what the interviewer wanted. While no one declined to be interviewed out of hand, some refused to discuss some types of data. Others put off appointments repeatedly and then were finally abandoned.

Work on Hawai'i was very useful in terms of contacts developed and information received, but after a rather large number of trips, it still remained spotty in coverage due to the difficulty in setting up interviews. Several trips developed no new entries in the data catalog for a range of reasons from

people being called out of town to lack of data.

However, the intensive interviewing schedule did have its advantages. Word of our work preceded us and on a couple of occasions, interviewees stated "I wondered if (when) you would find me". Some interviewees had very little to offer, but just wanted to talk about what was going on relative to data, systems, GIS and photography on the other islands.

This latter characteristic of most interviews, began the discussions of the need for data dissemination as the survey progressed. In these first intensive months working the Big Island, the concept of the project began to change. We discovered an almost profound ignorance of digital data and new federal products.

In the case of the Geology Department in Hilo, the interviewer ended up arranging for more data to be provided to them than received from them. That memorable two-hour round table discussion with part of the department changed the direction of many of the following interviews and the focus of the project.

For most data users, interim and timely information was more useful than complete metadata. At this point the Data Catalog Project probably should have been totally recast in terms of potential benefit to all of the state's players but the charge of Act 160 could only be bent so much.

*Project Expansion and
Newsletter Discussions:*

After the term of the first contract expired, it was obvious that the Data Catalog Project was badly under-funded given the accepted methodology. In response, the contract was expanded and broadened to allow a better sample of the data, but even with project expansion, it was becoming obvious that all the data in the state could not be found and surveyed with such a modest effort.

At this time, discussions also began about the very pressing need for general data dissemination about many of the issues that were being raised repeatedly, questions about digital data, progress on GIS, and dozens of other areas of broad interest. We also found many had an interest in the historical data available, especially aerial photographs, and very broad concerns about data accuracy and reliability.

When the pattern of slow field interview responses per time spent started to develop, three other interviewers were hired as well as a data entry person. After training, one of the interviewers covered Moloka'i, one started to work Maui, and one concentrated on Kaua'i.

The interviewers were variable in quality. The Kaua'i interviewer was excellent, displaying a tenacity for teasing out data from a wide range of sources including interviewees who did not want to be interviewed. One county agency person started out refusing to be interviewed. He ended up being most productive in terms of data while warming to goals of the project and inviting the interviewer back to "talk story".

Because of the skill of the interviewer, the Kaua'i data is probably the best sample in terms of breadth and depth. This interviewer also worked some of the O`ahu sources in spare time after Kaua'i began to produce diminishing returns.

The data developed for Moloka'i by another interviewer was good, but sparse. Intuition leads us to think that some data were missed. The Maui interviewer proved ineffective and found that he did not handle the process well and resigned. Later another Maui interviewer was hired and produced much more data in a shorter period of time.

*The First issue of
geo-Data Hawai'i:*

At this time, a data entry person was trained and field interviews were largely turned over to staff members while a test newsletter format was developed and sample articles written. At first the newsletter, titled *geo-Data Hawai'i*, was designed to carry short articles on data that the survey developed and expanding on the details of a project and the agency involved in the project.

There were to be short educational articles and definitions and periodic reports on the status of agency work. After several test drafts and replacements of articles, the first issue appeared with a review of traditional map and photo data in Hawai'i and an editorial on the data in general. The first issue was sent to an abbreviated list developed from the Contact data base and other interested parties and was well received.

This encouraging response stimulated planning further issues and outlining a range of articles. While work continued on interviewing and data coding, Hurricane Iniki causes several changes in the execution of the process. The damage to Kaua'i stopped all interviewing there and precipitated a special *geo-Data* issue that detailed local and federal response to the hurricane though special aerial photographic flights by different agencies.

Later Fieldwork:

As more time was spent on collecting data and writing for *geo-Data*, field data collection slowed down and we began to face the issues of keeping the data current while

further interviews were continued. Some major players were re-interviewed to update their information. The intensive field time of the early phases of the project were never revived as the interviews seemed to be losing ground with less data per interview hour being produced.

Discussions increased on methods of distributing the data base without resolution while *geo-Data* progress also slowed in the face of the lengthy editorial time that it was taking to complete an issue and other commitments.

Staff discussions negotiated an end to interviewing for the Data Catalog with a specific sample to be covered before closing the survey. More work was done on Maui and major agencies in Honolulu were surveyed.

Changing Newsletter character:

geo-Data Hawai'i also began to evolve to more of a technical and educational review function. Article composition of the later issues became subjects of lengthy discussions and required increasing amounts of research and editorial work while the frequency of the issues decreased.

As the tenor of local spatial data analysis changed with accelerating GIS interest, articles changed even more. It was decided that the newsletter should attack the issue of map and data accuracy in a major way. This decision involved months of research and editing to encapsulate this topic in a newsletter format. At last, three issues were planned and defined and the first was completed.

Externalities:

Publication of this issue generated some interest in the general accuracy topic and research began for the followup issues. Trying to keep pace with the timelines of the data base, updating and editing catalog data continued. Slow progress was being made on maintaining the Data Catalog and the next issue of *geo-Data*, local budget cuts, and agency reorganization placed progress on hold. These externalities forced ending the contract without completely satisfactory closure or Survey completion.

Closing the Contract:

Approximately half of the articles prepared for the newsletter were not published due to editorial work, incomplete research, dated material, and more pressing material coming to the forefront. The need for the service provided by the newsletter remained. However, even as the project was still underway, different methods of information dissemination were being discussed.

A last newsletter is in press, one that harkens back to

the original proposed structure. It has one informational article and several short news articles relating to new products and futures.

At the present, supporters of the newsletter function are discussing developing an Internet / World Wide Web home page to continue to provide information with the added function of distributing non-proprietary local data to all interested parties. Freeware now on the Web allows reading of AutoCad and DXF files, so vector data should become a regular part of Web distribution.

The ambitiousness of attempting a complete data inventory with metadata information in an interview approach with two part-time interviewers was naive. The Data Catalog contains information on only an estimated quarter to third of the data extant in the state. About a third of the data is dated.

There is a continuing and accelerating need for Hawai'i metadata. To fulfill this need for a current catalog would take a staff of a half dozen people about three years to complete and a staff of one or two to keep it current and indexed after that.

This task is a worthwhile one, if the hundreds of people involved in federal metadata efforts and the federal data clearinghouse are any indication. While the project was undertaken in enthusiasm and high hopes, it was too ambitious. No matter how much optimism was present, the Survey, as originally funded, could not totally succeed for lack of staff and technical support.

Hawai'i Data Catalog:

As noted earlier, the Hawai'i Data Catalog is a sample of the data sets available in and about Hawai'i. Data sets included are historical, small compilations, large projects, and projects in progress. Some the information is dated, not so much obsolete as incomplete.

We estimate that in a broad sense, a good sample of the major work is represented with maybe a half to two-thirds of the major lines of work in the data base. If much of the project data were disaggregated into specific data sets, as should have been done, maybe only a quarter to a third of the data available is contained in the data base.

Some of the larger players in the data base provided their information in pre-compiled form, which greatly aided in the preparation and editing of the Data Catalog. The players

which did not have their data in this form consumed over 90% of the time devoted to the development of the Catalog.

Data developed and collected:

The data included in the data base are not geographically equally represented. It is difficult to be definitive about geographic representativeness since many agency's neighbor island data is transmitted to a Honolulu office where it may be missed or double counted.

The interviews on Kaua'i developed the most complete and detailed data set as well as the best sample of the range of data available. Hawai'i was next best surveyed but the Kona area was not sampled. Kaua'i was the only island that included a significant amount of private data.

Data omitted:

As is reasonable, many private firms were reticent about releasing information about their data holdings, especially archeological research firms. If their data were included in the Data Catalog, a decision was made to block it from public search, however, later we decided to avoid private industry, except the large utilities, and deal with them separately in a later phase if time allowed.

Maui and O'ahu data were at best, a sample with holes in some of the federal agencies, particular in the military. Due to time, all of the City and County agencies were not surveyed as well as some of the smaller state agencies. Some of the larger public and private groups need much more work for the size of their holdings. The Bishop Museum, State Library, State Archives, DLNR, University of Hawai'i Library, Hawaii Institute of Geophysics at the University of Hawai'i at Mānoa, State Surveyors Office and other large organizations almost need a separate project of their own.

No data were identified on Lāna'i but it is expected that some exists in the private sector. Some data were noted for Ni'ihau, but several informants requested that it not be recorded. A great deal of data exists for Kaho'olawe, but it is very hard to find in many cases, beyond some of the USGS cooperative data.

Some work in the private sector was done for public agencies and shows up in their lists, but it is estimated that these firms hold a large amount of proprietary and untapped data.

Data discarded:

The staff and administration struggled with the definition of "catalog" throughout the project. To be a true catalog, every sheet of a map series and every file in a data set should

be listed. This was far beyond the scope of the project but badly needs to be done. A graphic index needs to be published held on an easily available mini-GIS.

Some of the Catalog entries are single sheets, but the bulk of the data are for "projects". A working definition of a "data project" is an activity that produces a data set. Some entries encompass thousands of data elements and some few. Although USGS topographic quadrangles and orthophotoquads are the base for many projects listed, these series are not listed as projects because of their universality.

Some "projects" use common data sets as mentioned earlier and these duplicate data sets are omitted from all but the parent agencies listing, e.g. the ubiquitous TMK series. However, this omission criteria might eliminate useful data because of the value-added annotations and modifications made by many agencies.

*Problems with Agency
Data base:*

The proposed "Agency" data base that was to provide valuable cross-linking was abandoned for several reasons about halfway through the interview phase of the project. First, details about the physical base used by the agencies, firms, and public groups were very difficult to collect.

In addition, many interviewees did not know their physical base without checking their inventories. Collecting data on facilities often devolved into "spying" and taking mental notes during walk-throughs of agency premises. Some groups were proud of their hardware base, others embarrassed, and some did not want to divulge the information.

When this data base was re-thought, there was concern that public access to the Agency data base could become a marketing tool for salespeople that would result in unwanted sales approaches. If giving information about these agency characteristics would result in marketing use, we felt that the very open cooperation that we experienced early in the project would disappear.

Added to previous consideration was the extreme difficulty in keeping this data base current. In many large agencies, the hardware, software, and facilities base changes almost weekly with the purchase of new equipment and the surplussing or sale of older equipment.

In the private sector, the public knowledge of equipment inventory and facilities base was consider proprietary information that could be of value to competitors.

For these reasons, the Agency data base was aban-

done early in the project and collection of information relating to this aspect of spatial data in Hawai'i was discontinued. Much of the most useful elements of relating to equipment and format are already contained in the "Activity" data base.

*Maintenance of the
Contact data base:*

The "Contact" data base was very important in the early parts of the project and was updated and edited extensively. However, next to the Agency data base, it was the most difficult to keep current. As firms and agencies grow and reorganize, staff change positions and locations. Just keeping the phone numbers up to date was a major problem, at one time costing hours on the phone each month.

The Contact data base was maintained for the first two years of the project until the interviewing scope was set and the final sample of interviews defined. The data base remains useful as a record of the attempts made for interviews and for a starting point for further work, but is several years out of date. When choices had to be made about limited time, it was decided to not use potential interview and clerical time to updating this data base.

Activity data base:

The primary goal and product of the Hawai'i Data Catalog is the "Activity" Data Base. It contains detailed information about approximately 700 data bases that currently exist in and about Hawai'i's cultural and natural environment.

The following section outlines the data fields with definitions and an overview of the data contained in the data base. In some cases, the field's data was not as detailed as we expected it to be when the field was defined. Some fields contain redundant data, some contain estimates, and some had to be redefined on the fly while data were collected or while undergoing the final edit.

This section is designed to alert the user about the nature of data contained and to flag survey sections that have proved to be of limited value.

*Data Categories and
Content Overview:*

The Activity data base contains the detail of the specific projects of interest. Each project or logical project grouping has its own record. In some cases a record will exist for a single map. In other cases, a record exists for a collection. In other situations, depending upon the nature of the data, a record will encompass a functional group of similar mapping projects.

Agency/Firm: The data contained in the Agency/Firm field is sometimes abbreviated to the more commonly used form of an agency name. As we all know, the full administrative chain of command that is reflected in a formal name, is often abbreviated. For example, the variety of departments of Public Works that exist on each island are entered "Public Works (Maui County)" and the Department of Agriculture is entered alternatively as "Agriculture Department (US)" and "Agriculture Department (State)."

Division: self explanatory

Branch: self explanatory

Location: The island location of the office or agency where the data listing is maintained or stored. Although data may be collected on a neighbor island, it is often transferred to O`ahu for use. If the data is maintained in both offices, both may appear.

Project Director: If a project is administered by an individual other than a Division or Branch Director, this person is listed as Project Director. If the project is the result of a staff member assigned the task, the Division or Branch Director will be listed as the Project Director. If the person in charge of a project that produces data is unclear, the field is left blank.

Contact: The person who provided either the interview which developed the project record, or the person who conducted the project. In some cases the Project Director and the Contact may be the same. Where data were provided in pre-compiled form, the person who provided the interviewer the compilation is listed as the contact.

Phone: self explanatory

Client: The agency, person or firm who commissioned the data to generated.

Project: Brief description or title of the project that contains or is the focus of the mapping activity or collection of data.

Purpose: This field was intended to narrow the appropriate accuracy limits of to specific applications. However, in reality most contacts and other data providers were confused by the question, appearing not to know that data are usually matched to an intended application. Therefore, the interviewer usually made an estimate which contained less specificity than hoped would be provided by the contact.

Other Data: A class of thematic data is usually mapped with other supporting or related data. If these added classes have general interest as a data source, they were noted here. If this field is blank, it means that the data was mapped or held in a

file in isolation, as would be the case in many GIS layers.

Information date: Date that the information on a specific project was collected by interview. If the date of the interview was missing, the date of a project listing provided by an agency is substituted. If neither of these dates are clear, the data entry date is used.

Area: Island that was the subject of the project. State is used if multiple islands are involved in the project. If all islands are not in the project, this is noted in the extent field.

1:50,000 quad: This field was intended to be a searching field if a specific area was desired. We chose the 1:50,000 quad since it was narrower than a whole island, but larger than a 1:24,000 quad. Since this is the only topographic series that contains the current, correct UTM system, we thought this was appropriate. However, this defining series turned out to be an unfortunate choice for reference since the 1:50,000 series was discontinued by Federal budget reductions.

Judicial District: An alternate searching field for areas that might be more familiar to some users and one that is more closely geographically linked to island topography.

Extent of coverage: More detailed description of area covered in the project scope. Usually it is a qualifier that limits one of the earlier fields. It was intended that this field would be replaced with a graphic index map at some time in the future.

Index: A note on the existence of a map or spatial index plot for the project's data collection units.

Completion: This field was originally intended to indicate whether a project was complete, on-going, or in-progress. However, these categories are hard to define in some cases such as data in a GIS. Some types of information can be complete but added to, as changes occur, for example the parcel base. However, this does not really fit on-going or in-progress categories, and should be redefined.

Revision date: Date of a planned revision that will be reflected in the currently available data.

Completion date: Date of completion of the original project. Again in terms of the realities of GIS and other digital projects, this field does not have much meaning. It is more appropriate for use with traditional mapped data and aerial photography.

Data source: The information source(s) that was used for a map or data base. It turned out that this field was particularly important for estimating some of the parameters related to data characteristics. Often when a contact was unfamiliar with the concepts related to spatial resolution, it would be possible

to limit resolution based on a knowledge of source data. For example, if a vegetation map is based on a topographic map, the resolution of the data cannot be better than that of the topographic map.

Data date: This field makes a distinction of data date versus completion date, although sometimes they are the same. In future work, we think that this field could be substituted for completion date unless completion data is redefined.

Scale: Scale of a hardcopy entry such as a traditional map or aerial photograph. This was used as an indicator of appropriate use. In the case of digital data, which is essentially scaleless, a "n/a" was entered. If data were digitized from a 1:24,000 quadrangle, a "n/a" was still entered, because the nature of the digitizing could substantially change the spatial resolution characteristics of the original 1:24,000 quad and its appropriate use.

Spatial resolution: This field represents the probable precision met in collection the data or mapping and digitizing it. This turned out to be a particularly difficult field to complete because most users had no idea of their data's precision. In most cases, the data was coded with an estimate of the resolution based on the source of the data. For example, field surveyed data is always better than a meter, while resource grade GPS is about 40 meters. Data based on topographic maps was estimated, based on the probable accuracy of the topographic map from National Map Accuracy Standards and reduced according to compilation method.

Information Classes: This field adds information to the Other information field. If the map is a vegetation map, the number of class of vegetation is noted. In the case of a single theme map such as Legislative districts, simply the boundaries are listed as the information classes.

Reliability: General consistency of the data over the project. This field was a value judgement based on the completeness of the source compilation, or a contact's estimate of how good the data was. It was hoped that this field would be a searchable field, but in practice it represents a quite weak estimate.

Accuracy: A verbal statement of assessed overall data quality. Intended to be a search field more general than spatial resolution. This field is probably more useful than the reliability field because it is based upon compilation method and source information, but still a secondary level estimate.

Digital format: This field was intended to reflect the specific data storage format such as ASCII, TIFF, DXF, DLG3, etc.

However, most current digital data are stored in a proprietary software format such as AutoCad or Arc/Info. Therefore, this field can contain a software name which is not completely appropriate to the field definition.

Storage Media: The storage or distribution media. eg. optical disk, 3 1/2 " high density floppy disk, 9 track tape, etc.

Hardcopy Type: The final form of hard copy materials, eg. a color map, a plotted graph, a tabular printout.

Hardcopy media: The hard copy base media, eg. scale-stable 4 mil mylar, paper printed copy, hand drawn vellum.

Hardcopy size: In the great majority of interviews, the interviewer did not actually see the hardcopies or an index. Therefore this category and the next can be an estimate. Most contacts did not know how many sheets there were in a collection and there was not time to carefully count and inspect the holdings.

Number of sheets: Number of sheets in the coverage.

Availability: Status of the data with respect to loan, copying, purchase, or distribution.

Copyright status: self explanatory

Data published: self explanatory

Cost of data: self explanatory

Notes: General or specific important information not detailed above.

Hawai'i Metadata:

The Activity data base is structured in a way that is similar to the metadata standard established by the federal government. However, this standard is far more detailed, one that could not be completed in an interview approach. While this data base may change the way we think, a great deal of additional information must be added. The federal metadata template requires very detailed information on the originator, naming, and publication status of the data.

In addition, spatial domain information is important and each sheet must include bounding coordinates in addition to detailed use limitations which are separate from detail on attributes, consistency, completeness and positional accuracy. Detailed information on the characteristics of the map projection used are also included, as well as very detailed access data. The digital information included in the metadata may go into detail about digital access to the data via modem characteristics, ftp addresses and further information relating to cost and security.

While it is unlikely that Hawai'i generated data will

need this level of detail soon, the 55 page "Content Standards for Spatial Metadata" from the Federal Geographic Data Committee should be a primary reference in defining further data classification and heritage forms.

**Methodological
Review:**

Field work: We remain convinced that the field interview method that was employed in this study was the best method for collecting information about data, especially the small and obscure data sets. Large data sets such as are present in the State GIS are more efficiently entered from physical lists, but in most of these data listings, followup interviews are necessary to add details that are missing.

If we were able to make one suggestion to firms and agencies, it would be to emphasize the importance and usefulness of maintaining a complete listing and index of their data and its characteristics. Such listings would greatly expedite a similar Data Catalog in the future, but also would help maintain internal collections. On more than one occasion, we found data in an agency that had simply been forgotten.

In the future, we think that several levels of interviews would be useful to make the metadata compilation more complete. We would suggest a first interview to get the listing of data with whatever estimates of metadata characteristics can be made by discussions and examining sample maps. A first stage interview similar to what was actually done in this project. When this listing is coded, it should be sent to the respondent for review and then an additional meeting should be made to edit the listings.

In most cases, the second meeting should be adequate in order not to interrupt too much. However, in some cases, several meetings might be necessary. For this reason, we feel it is important that this type of survey be conducted by regular staff who could make periodic visits to update data and review the status of new work.

It is most important to be prepared to give contacts information as well as request data. This interchange of information is important to gain the full cooperation of contacts. This cooperation can be best gained by being willing to sit and chat with the contact and adjust to their schedule and method of interaction.

A long-term schedule of data collection must be expected. We judge that interview style of the data collector is crucial. The interviewer must be seen as a colleague that is genuinely interested in the contact's job and responsibilities, one who wants to learn and share. The interviewer should have a professional knowledge of the contacts data and agency so that talking informally can be profitable.

It would be a mistake to send an untrained student intern, not familiar with Hawai'i government and business practice into a local county agency. What must be emphasized is that the interview should produce information from which data can be derived. To make an appointment simply looking for data will almost always miss a great deal because a contact may not consider much of their work is producing data.

It is also most important to be able to look at the data, whether in hardcopy form or on a monitor. A great deal of information about data classes, resolution, and data quality can be gleaned from viewing the data by a trained interviewer. However most important, viewing data almost always produces questions that provide additional information. Contacts are almost always proud of their work and discussions over the actual data can spawn rather far-reaching discussions of appropriate applications.

If at all possible, when confidence is gained, the interviewer should seek permission to browse map files and examine aerial photographs for additional details, both omitted or forgotten. Most such files contain an index sheet that can be useful for filling many format blanks. It is far less likely that an interviewer will be left to play with data on a computer, so approaches to exploring this type of data requires a more "show and tell" approach.

The flash of computer systems and the agencies pride associated with them sometimes can be exploited. With these centerpieces, old traditional data, even if is the basis of the digital forms, are likely considered to be unimportant and forgotten. In the case of the Hawai'i County GIS, the source data provided far more information than the screen data after both were seen.

After some of the excitement of the original approaches to contacts dulled, persistence becomes more important and the interviewer needs to cultivate an element of excitement in their voice. Even when being shown repetitive

and low quality data, the interviewer must be excited in each "new" find. It is important not to concentrate too many interviews into a day since it is hard to remain fresh. At least every interview day should be half devoted to data coding.

Before any firm or agency is approached, it is important to know what they do and whatever possible about their divisions, branches, data, facilities, and chains of command. When the interviewer can ask intelligent questions about an office's operations, interest in the interview will usually be quickly established.

The hardest part of efficient data collection is scheduling appointments that are long enough to be useful at a time when people have time to talk. We found that some, if not most, groups preferred early morning interviews. Contacts were fresh and had not yet got involved in a day's work. We let the contact determine the length of the discussion, when contacts loose interest or have conflicting phone calls, leave and come back another day. Avoid all late afternoon interviews. This time slot always seems to tinged with identifying the interviewer as an impediment to the desire to get out of the office early.

The largest mistake made was giving up too early on scheduling appoints and followups. A great deal of data was lost by skipping people seemingly too busy to talk. As much as maintaining the important excitement in an interview, is a sense that the interviewer thinks every interview is important. Repeat calls may be necessary to convey this sense, but if persistent and courteous enough, data can usually be collected.

Data base: It is important to allow a minimum period of time between coding and entering data. When information is fresh in the mind and images of the maps are still in the short term memory, there is less temptation to enter information that "should" be rather than what was actually seen.

Before actually entering data in the data base, we would suggest an interim step of entering data on a matrix where boxes can be checked rather than directly entering a value judgment. This intermediate step can provide a type of visual scaling of some classes of information that may allow an adjustment of values on the matrix as a whole before final coding.

In the future, technology may make a very large difference in data collection. Since most current data bases support

graphic information we would suggest that an interviewer take a hand scanner and laptop to appointments. This equipment will allow part of the map or photograph or its legend scan recorded for future reference and coding help. This is far more difficult with digital information, but wherever possible a small portion of a printout might be requested for off-site scanning or a digital camera shot of the monitor screen would be most useful in a similar situation.

Conclusions: In general, the methodology used in the Catalog project was valid, however, the data base should have been more detailed to reflect current metadata directions. The largest problem was time. The amount of data that is present in Hawai'i was just too large to be properly surveyed in the time available. To conduct this survey properly requires a large staff which should be part of larger clearinghouse and service bureau function.

The project was useful and produced valuable data, but it simply was too tightly defined to produce a data base that will serve all needs.

Epilog:

The experiment that was the Hawai'i Data Catalog Project taught us an number of things beyond the data surveyed and the establishment of the beginnings of a communication network through *geo-Data Hawai'i*. There are many more data sets in existence than first estimated. Some of these multi-faceted data sets provide a knowledge about our islands that few use or imagine. In addition to the richness of data that exists, there is a general ignorance of the this data. The data's location is only poorly known, its attributes are generally unknown or known to only a very few.

Very few people know how to fully access the available data. Even fewer have developed a vision about what its imaginative use could do to improve our knowledge and the quality of decisions we make about our future. We are in a time where more and more data are necessary to make informed and difficult decisions about our society and environment. Access to current, accurate data seems to be diminishing, even given the powerful Geographic Information Systems that are being put in place. In this era, often loosely called the Information Age with its exploding data, progress in putting good information in the right hands seems does not seem to be

keep pace with the rate of generation of new data.

*Changing Spatial Data
Environment:*

At about the same time that the Data Catalog Project was discovering the magnitude of the problem of finding and categorizing data, other factors were occurring causing changes to the development and use of spatial data. The seriousness of the fight to control the national budget deficit which affected all levels of federal services and agencies and local budget cuts has forced re-thinking of mapping and data priorities both in Washington and Honolulu. Before the current series of assaults on the US Geological Survey, the provider of much of our base data, other changes were in the wind. The Reagan administration privatized the earth-focused resource satellite program LANDSAT. While the taxpayer paid for launches and many other elements of the program, the data products from the program became much more costly.

Under the privatized LANDSAT program, satellite images are licensed to one group or person for a single application -- no more of sharing public domain images. The cost of an 8 x 8" black and white photographic print from a single scene jumped from \$1.25 before privatization to \$200 after. A \$200 digital scene increased in price accordingly.

The effect was to largely stop public use of the imagery. The effects of this program were felt in the establishment of the better French SPOT hardware and imagery program and after several years, the beginning of de-privatization.

The same administration formed an attack on the National Ocean Survey, cutting staff from 450 to 100 in that NOAA office in the hope that private industry would take over nautical chart production. That plan was never successful but the NOS was left seriously weakened.

*Forced readjustment of
USGS mapping and data
priorities.*

With the rush by both branches of Congress to eliminate the budget deficit, all manner of plans that affect the data infrastructure of the country emerged. With Administration pressure over the last several years, the priorities of the National Mapping Division of the US Geological Survey have resembled a tennis ball that is off the court as much as in play.

As the USGS was settling into a major development of digital mapping data nation-wide, and completing the 1:24,000 quads, they joined into cooperative work leading to a nation-wide digital data base for the 1990 census. The TIGER data base was not finished in time for the census at USGS standards because of time and the fact that the base 1:24,000

mapping was not completed.

After that time one priority after another has dropped by the wayside. First the 1:24,000 quad priority was eliminated in favor of orthophotoquad production. Soon analog orthophotoquads were dropped in favor of digital orthophotoquads with nation-wide coverage planned by the year 2000. Map production priority moved to the cheaper 1:100,000 series but in particular, their digital counter parts which have proved very successful.

More digital products are becoming available at the expense of revision of quads and new mapping. To counter some of the loud criticism of the quad program priorities, the Digital Raster Graphic Program was announced - scanned versions of paper quads. This program became the new first national priority, even given the still existing coverage gaps at 1:24,000. Targeting to be finish the DRGs by 1998, in a rush to show Congress completion of something, the existence of the agency itself came under assault in several waves of budget cutting.

In the 1960s the USGS began talking about the long term transition to a digital product. The popularity of the DLGs and DEMs speeded the transition. As the digital revolution increased in speed and the budgets became tighter, short-cuts were taken. With new products introduced, there was a good deal of disarray in the agency.

Current budget cuts have probably accelerated the transition from a paper map based agency to a digital data focussed agency. With this transition, many quality and accuracy concerns have fallen by the wayside. The National Map Accuracy Standards have practically been discarded and there have been official discussions of relaxing the already weak standards in the guise of production efficiency.

At the present time digital products are being dropped. the DLG-E program has become strangely silent, the Hypsography DLG has been abandoned. All new "mapping" projects now need local money or the cooperation of other federal agencies to move forward. There is also a move afoot to make what is left of the USGS an expediting or organizing agency, giving priority to helping national and local groups in their GIS and mapping efforts.

*Localization of Data
Production and Editing:*

The outcome of shifting emphasis of data production to the local scene is too hard to predict. However, many

Western States are taking over their own mapping and ortho-photo production, sometimes meeting USGS standards and sometimes not.

When questioned about updating current products, the USGS shrugs their shoulders. They say that in the future they will be making revisions and updating products in some digital form. No one seems to know what that means and how it will be accomplished. Will the roads layer of the DLGs be updated when new photography becomes available or will the USGS integrate locally produced data, created to eroding standards? Will local data created with local priorities and local funding become the standard. Probably the latter.

*Data Distribution and the
World Wide Web:*

The USGS still sells the 1:24,000 DLG information at a pittance and only distributes it on 9 track magnetic tape. It is too early to know what will happen to the 1:24,000 DOQ and DRG CD-ROM distribution prices. These files are large and inefficient to transfer via Internet over a modem. However, many data sets are now available free over the Internet. All of the 1:250,000 DEMs for the nation are on the Internet. Also, all of the available 1:100,000 and 1:2,000,000 DLGs are available on the Web. In time, the 1:24,000s will also be there.

Updating the data bases is the issue. In Honolulu, the State GIS has updated the DLGs for their own use. The Water Resources Branch of the USGS has modified them as well. University groups have also updated them along with a number of private agencies.

Will these revisions find their way back to the USGS for general distribution? Unlikely. If the USGS were to set themselves up as a clearinghouse, it would take all of their time checking data and modifying their metadata files. With the transfer of easy editing to the local scene by creating DLGs, the situation is likely to become one of disorganized anarchy.

Local agencies are already taking steps to put their own versions of the DLGs on the Web for public distribution. The University is about to put NAD83 versions of the DLGs on the Web for O`ahu and other such distributions are in progress.

This anarchy will in some ways put topographic base mapping back to the days of the 1930s when local agencies did their own topographic mapping to meet their own needs. However, the difference is that standards will be developed locally to meet agency needs and probably well hidden by the

ease of digital production, editing, and distribution.

*Data Libraries and
Service Bureaus:*

If this situation is not to turn into a total anarchy, local data libraries responsible for standards and distribution will have to be created. It will become necessary to form these clearinghouses in order to avoid an ever-increasing duplicating of effort and a total dissipation of accuracy standards. The State Geographer/Cartographer position and office that was formed in the late 1980s in Hawai'i should have been in charge of this data cataloging, verification, and distribution. This did not happen and given the current budget situation will probably not happen soon. And a great deal of disarray and confusion will have to be endured until it does happen. We think it is a question of when it will happen, either in the State government where it belongs, or in private industry. Already there are many mainland firms who collect, reformat and sell USGS data - often badly out of date - at substantially increased prices.

The State desperately needs a data clearinghouse to distribute data, raw and value added to the public, private industry and the other agencies. There have already been moves made to force the agencies to provide data to anyone who asks under public disclosure and other precedents. This direction is lead by ESRI and other GIS software distributors to facilitate the sale of more software.

For good or bad, ignoring liability issues, this trend to public access for most data will continue. And in our opinion, a public agency should be in charge to insure some official stamp on standards. Private industry should be a part of the process, but continued adherence to known accuracy standards must be a part of the heritage that was started by federal agencies.

Data Accuracy:

Data accuracy is not just a 'bother' to be abandoned as soon as possible to speed production. With advent of better collection and analysis procedures, spatial data will increase in quality. With the exception of a few private and public agencies charged with the production of data, data production should remain as an in-house activity. Those who know the topics that they are mapping and codifying, generally know it best, and can correctly demand and specify its characteristics. However, to bury data for in-house use only is a tremendous waste of money. Funds that could be used for profit, support of functions, and the public good. Data should be collected

according to published standards and for the most part be made available to support better decision making state-wide.

Data in the 2000s:

By the new century, much more data will be available and easily accessible via electronic means. There is a great danger that these data will be of variable quality if no one is in charge of verifying production and is responsible for maintaining accuracy standards. Just a few years ago, it was a major effort for an individual to acquire and analyze detailed census data and link it to local data. Today home computers and the Web put this capability in the hands of anyone who needs it, and is willing to take the time to learn a small amount of software.

Private groups and research agencies are mapping data at unheard of rates, other groups are checking earlier data for quality so that it can be used with more confidence. Agencies can now make decisions with more authority than in the past, and private groups have access to the same data to check their work. Old data will not be discarded, but newer data will give historic data a context. Informed decisions will be made when it is in the public good to discard poor data and negate decisions made from it.

Unfortunately it is difficult to edit digital data without extensive field work and new data will be created with ever more field work. Field studies, data creation, and data maintenance will go hand in hand. Data will become better and more extensive as a result. More sophisticated analyses will be possible and data will be more easily accessible, but what about the form of the data.

Maps in the 2000s:

As anticipated by many, paper maps will become more of a specialty item than in the past. However, their need will not evaporate. When pressed on this issue, Jack Dangermond, President of ESRI will talk at length about wristwatch computers and high resolution displays with radio links to data bases. But any who has tried to see the broad pattern of the landscape on a small monitor remain skeptical.

A compromise will be service bureaus who will tailor "pizza-maps" with a range of toppings, scaled to the needs of the field person. These bureaus will printout hard copies on demand or transmit them to your home computer for personal printouts on color photo-realistic printers that are now available to everyone.

Atlases are now prepared in this form, but the issue of

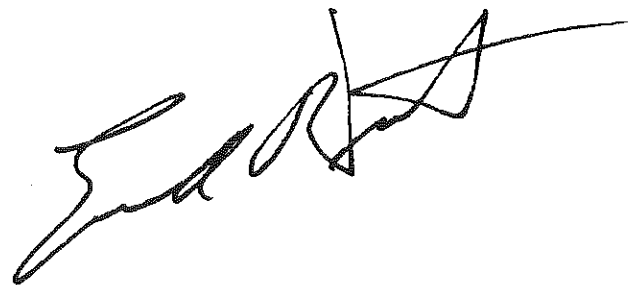
data currency remains. Who will be in charge of the data bases and their accuracy. What will be the charge for such a service, will hikers revert to old paper maps xeroxed many times and annotated from years of field use.

We are likely to see unexpected changes in the next 5, 10 and 20 years. No one will guess beyond that. Where will virtual reality engines come into the long standing need to image and study process in the environment. Most of us will not live long enough to see the changes in spatial data and mapping that 2020 will bring.

However, the need for mapping and data clearing-houses will grow and need to be structured in a way that will insure quality and public access. Hawai'i and all states and local areas need to have access to a catalog of data with detailed information on quality and data characteristics. We will also need to make provisions for accessing the data. Agencies and individuals will continue be able to have economical access to hard copies that we can take into the field to study and edit.

More and more people will need the ability to take information into a central agency who will update the central data bases. The State needs to develop such an office of data maintenance and distribution. The State GIS is not such an agency, its charges are for analysis and planning, not general use.

It is hoped that little projects such as the Hawai'i Data Catalog, and large projects such as the Federal Data Clearing-house will point the way. If little of the future can be seen clearly in these times of transition, one point is obvious. Spatial and other data will more and more become the responsibility of local agencies and planning should by underway to shoulder this very important responsibility.

A handwritten signature in black ink, appearing to be 'E. A. ...', written in a cursive style.