OGIC Document Number XX

GIS and Confidentiality

Joint Task Forces from the:
Administrative Boundaries &
Oregon Cultural and Demographics
Framework Implementation Teams

December 2002

Oregon Administrative Boundaries & Cultural/Demographic Framework Implementation Teams
Department of Administrative Services – Department of Agriculture
Department of Parks and Recreation – Department of Transportation – Employment Department

CONTENTS

Summary

Confidentiality of data and information is an issue affecting many federal, state, and local governments as they attempt to use GIS as a tool in making informed choices leading to good policy and decisions. Government is buffeted by the competing demands for public information and the need to protect the confidentiality of information.

Confidentiality restrictions lead to duplication of effort and increased government cost when all governmental budgets are under intense pressure for increased efficiencies. Governmental entities are collecting information and creating databases already collected and created by other entities simply because one or all cannot share their information. Government has the ability to analyze and present data and information in ways that were not thought of 15 years ago. Yet, most, if not all statutes governing confidentiality and the sharing of information were written before decision-making tools such as GIS, Modeling, Impact Analysis, and Performance Measurement were developed.

GIS, because of its ability to combine computer based mapping, database management systems, and powerful statistical and data analysis systems, complicates the confidentiality issue even more. Is a dot on a map confidential information? Scale, statistics, data categorization, mapping methods, and aggregation methods all can impact whether GIS data violates confidentiality requirements.

Currently, there are three approaches to GIS-related confidentiality in state government: Share everything; share only with specific partners for specific purposes; and share nothing. Some agencies have websites where the public can view detailed information such as property tax assessments, water rights, or hazardous waste sites. Other agencies post only aggregate or summarized data.

Principle should drive policy. Where possible, information should be shared between government entities and with the public. Statutes and policies that prevent sharing of information in cases where harm to the individual is not an issue, or where the right to confidentiality can still be protected, should be examined and if necessary, modified or eliminated.

Confidentiality and the sharing of information must be addressed as more agencies begin to use GIS and other tools to make policy and decisions. Failure to do so will cause decisions to continue being made in a vacuum.

Confidentiality and Geographic Information Systems

This paper comes as a result of confidentiality and data sharing discussions at the Geographic Program Lead (GPL) and GIS Policy Advisory Council (PAC) levels.

Confidentiality is an important consideration related to many activities involving collection and dissemination of state government data. Confidentiality can profoundly impact an agency's ability to release data for analysis and decision-making. Federal and state agencies in particular need to protect confidentiality, but are simultaneously obliged to report data to the public (Karr, et al 2001). This creates a complex problem for the data steward. The primary key is to balance (CMRC, 2000):

- Right to know versus right to privacy;
- Technological developments for disseminating information;
- Specific spatial issues related to scale;

Geographic Information Systems (GIS)

GIS allow the display and manipulation of spatial data to be easily completed. Many of the potential data sets that can be used in a GIS may have confidential data. In some cases, mapping this data, or cross tabulating it by unique administrative or jurisdictional boundaries, is important for resource, economic, and organizational planning and management. Therefore it is important to be able to use, display and document data while not violating individual's rights and privacy.

A large amount of confidential data currently exists in state agencies in Oregon. Some is related to specific human activities such as employment, taxes, mental health, welfare programs, and other social programs. Having access to this data is becoming more critical to making effective data driven decisions and reducing costs in governmental organizations and services. These costs are associated with building and maintaining databases, replicating efforts, and using objective information to make reasoned and defendable decisions.

In addition to being a mapping tool, GIS has a powerful set of tools for manipulating and analyzing data. For instance, through a simple process of overlaying data, one attribute can be transferred from one database to another. Examples of GIS analysis for human resource agencies include adding legislative and senate districts to data files for statistical summaries, determining distances and areal extents of resources, and modeling future growth. Other human related data includes demographic and cultural data (i.e. regarding some aspect of human activity). Historical information (cemeteries, anthropological sites, historic sites, etc) in many instances is confidential data. Maintaining the confidentiality of the data is critical so that a culturally important feature is not disturbed. However it is important to know the location of the features for planning purposes. Examples include using archeological information in the design and planning of roads.

There is also a large amount of confidential data related to the physical environment including locations of threatened and endangered species, mineral resources, pesticide applications, etc. For example the location of nesting site for endangered species is important for preservation of the species. Many critical facilities and infrastructure databases are also partially confidential for safety and security reasons.

Confidentiality Background

Two major sets of laws exist nationally and in Oregon related to analyzing and disseminating data. The first maintains confidentiality and protects an individual's rights or a resource. The second relates to the public dispersal of information. In many cases, there is no clear-cut dividing line between these opposing mandates (i.e. in the case of public safety an individual's rights may be negated). GIS methods have historically been used extensively for analyzing and displaying extremely sensitive data that are regulated by rigorous federal laws such as the Endangered Species Act. Other databases in Oregon currently have public web based GIS/mapping interfaces for data retrievals. These include individual property values and taxes, public and domestic wells, irrigation locations and water rights, water quality sampling sites, toxic waste and hazardous chemicals. The relationship between protecting confidentiality and data sharing in GIS is also difficult because GIS can use a map to display information, and in many cases a map may violate the intent of statues that were legislated to protect the individual's right to privacy and confidentiality.

Technological advances in GIS have made mapped confidential data more readily available. Recent advances in geo-coding methodologies allow address based data to be mapped. Other data sets such as detailed address based road coverages and detailed zip+4 extension centroids, allow large historical databases to be mapped with a high amount of geographic precision. Additional advances in Global Position Systems (GPS) also allow higher resolution geographic data collection and mapping.

Components of Confidentiality

There are several major components of confidentiality with respect to GIS data. These include the reasons for protecting identities; the methods that are used to meet this goal, including summarizing and aggregating spatial data; and potential methods for data sharing.

One of the primary goals of confidentiality is to protect privacy by not allowing an individual observation (employer, employee, benefit recipient, archeological site, spotted owl, etc) to be identifiable. Maps, specifically point maps, define unique locations that may be identifiable. However, identification of a particular point is most often dependent on an individual's knowledge, experience, and a host of other factors such as exposure (time), scale (distance), physical abilities (vision/hearing), etc.

GIS data are not usually based on specific individuals but rather unique geographic locations such as addresses, latitude/longitude, Public Land Survey (township, range

and sections), zip codes and/or watershed. Most all of this data is publicly available from local sources. For instance, addresses are available from several potential sources including the county planning department, phone books, private marketing organizations and local county tax assessors.

Database attributes used in a GIS are expected to have various levels of confidentiality. For instance the location of the recipient of a state program (such as welfare) may be less sensitive than the amount of the program received. Correspondingly the location of an individual's house and property tax (currently public record), is less sensitive than the individual's personal income and taxes.

Scale

One of the most important components of a map and/or GIS database is the scale of the data display. This is a expressed as a fraction of the map distance to the real world distance. A map scale of 1:63,560 means that one inch on the map represents 63,560 inches (or 1 mile) on the ground. As the ratio get smaller (>1:250,000), the map covers a larger areal extent. Large-scale maps (such as 1:24,000) usually have a smaller geographic area represented.

Whether an individual is identifiable or recognizable is directly dependent on the map scale. When large areal extents are mapped (i.e. small-scale maps), it is difficult to identify an object without prior knowledge and/or auxiliary information.

All mapped data is a representation of data at lower spatial resolutions. In some cases, point data, such as spotted owl breeding areas are mapped at a larger area (for example one square mile, or larger) that will not reveal the specific location of the critical resource. The map symbol is also impacted by the scale. On small-scale maps, points correspond to a large geographic area, whereas on large-scale detailed maps, points define more precise locations. Table 1 presents the map scale and the real world size of a fine line and small point symbol. This table demonstrates that on an 8 ½" x 11" map of Oregon a point is plus/minus 2 miles in geographic precision.

Table 1. Map scale and map symbol size.

Scale	Example (8.5 x 11)	Line Width	Symbol Area
1:2,500	Tax Map/Plat	3 feet	12 feet
1:24,000	Topographic map	33 feet	120 feet
1:250,000	County/City map	330 feet	1200 feet
1:2,500,00	State map	3300 Feet	12000 feet

Statistics/Tabulation

In many cases, the mapped data is summarized or aggregated, and is not presented as a unique enumeration or point value associated with an individual. This data

corresponds to the statistics of a distribution. A statistic, by nature, does not reveal specific information related to an individual. The statistics that are often mapped include categorizing quantitative data. Quantitative data (ratio and interval) is typically classed into equal intervals (e.g. percentiles), standard deviations, natural breaks and other methods for partitioning data. When data is classed and mapped, the actual values cannot be determined (only potential ranges). If done correctly, this method allows the communication of information while still protecting confidentiality.

Data Categorization

Data categorization methods are typically used in/with GIS technology for analyzing, synthesizing, and integrating data.

There are two primary reasons for data categorization: 1) for easier interpretation and, 2) to protect specific disclosure.

Mapping Methods

There are numerous methods that can be used to map data. The three most common types of data mapped are points, lines or polygons.

Point mapping almost always presents a unique geographic location. However, as noted before, scale will impact this.

Line mapping can represent unique features such as a road but can also represent magnitudes of a quantitative distribution.

Polygon mapping typically involves aggregating data into administrative or jurisdictional boundaries and/or identifying unique material types (i.e. soils).

It is common to aggregate data to administrative boundaries (city limits, counties, neighborhoods, zip codes, etc) and to map and tabulate the data. This approach has advantages and disadvantages. The major advantage is that computer programs can be created and applied between databases. Several disadvantages are also associated with aggregating data. One is that the rules can be difficult to apply. Another is that many administrative boundaries change over time and the program or run becomes obsolete or requires change itself.

Spatial Aggregation Mapping Methods

Several methodologies can be implemented in a GIS that allows mapping of n-level data. Figure 2 presents a methodology by which data can be aggregated into cells that do not violate confidentiality by ensuring a minimum sample size. The aggregation unit (in this example represented by the box) can increase in areal extent to always maintain a minimum sample size (n=> 3). This hierarchal method will allow the distribution to be mapped and released while not violating confidentiality.

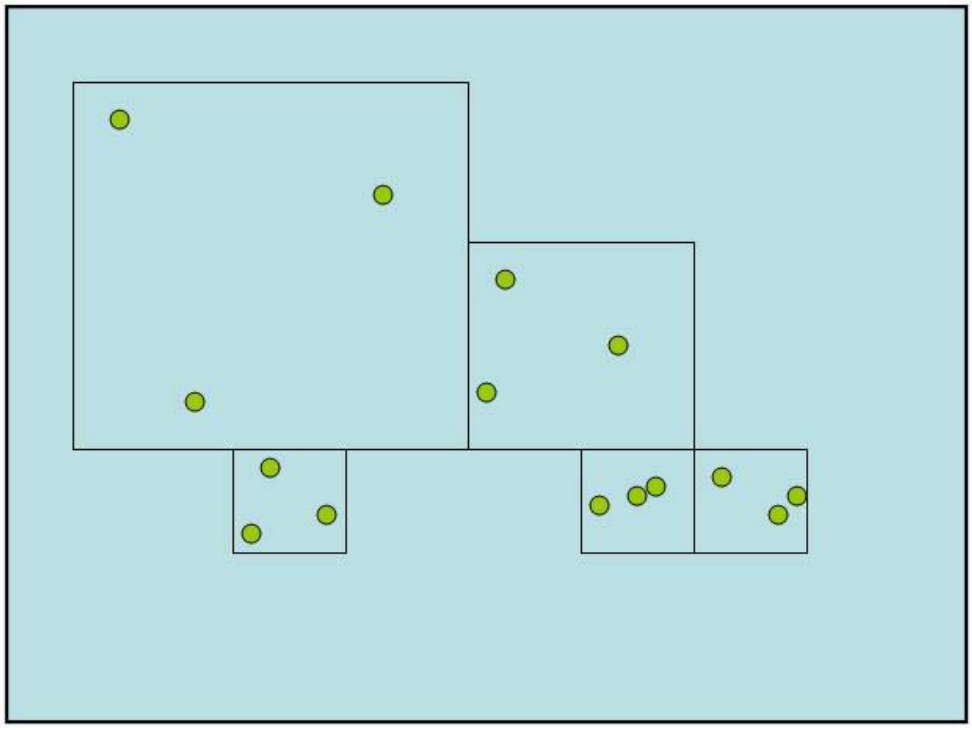

Figure 2. Spatial aggregation of point data in nested boxes.

Other methods can also be used to aggregate spatial data. Multivariate statistical methods such as Cluster Analysis, Principle Component Analysis and Factor Analysis are commonly used to reduce dimensionality of complex data. Mapping data in this manner is common for complex databases and has been used extensively in ecological mapping. Interpolation methods are also used to extrapolate site specific data to larger areal extents. Other neighborhood based GIS functions (nearest neighbor, buffers, thiessen polygons, etc) can also do similar analyses.

Current Data Sharing Strategies

Wartell and McEwen (2001) have noted that data sharing issues are mostly not technical problems, but are rather due to politics and personalities. Setting up strong, logical methods for cooperative data exchanges is critical to this endeavor.

Some states have set up cooperative data sharing arrangements specifically related to GIS activities, in order to minimize costs among and between state programs and to ensure high quality data that is transferable between applications.

Other states have data sharing strategies between agencies (such as Employment, Transportation, Human Services, etc). Maryland also has an established interagency sharing agreement that specifically outlines data flow to promote fast and effective geo-coding between agencies. Several agencies (Human Resources, Justice, etc) share in building a master database and thereby reduce costs between programs.

Currently in Oregon State Government, there are three primary models of data sharing. One is to publicly share and distribute most data. This approach is common in natural resource-based agencies and is typically done through Internet based interfaces. The second model is to share a limited amount of data based on specific requests and applications. The Employment Department is an example of this approach. The third is not to share any data with other organizations. The pesticide reporting application is an example of this approach.

Other Data Sources

In many instances, there are numerous other alternative data sources, which can be displayed or mapped instead of using confidential data. For social and cultural databases, this includes the Atlas of Oregon, and/or digital data from the telephone book. Other data can be purchased from commercial vendors who provide data similar to agency based data sources. An example would be InfoUSA, which provides geo-codable data with estimates of employment, sales and industry type. These data sets are typically available at a fee and represent estimates of the state collected data. Similar potential databases are being prepared for environmental data (Willamette River Initiative) by academic and nonprofit organizations.

Federal Laws

There are two primary federal laws that are related to the topic of confidentiality. One is the Freedom of Information Act of 1966 (USC 5 USC Sec. 552) which ensures public access to government collected data. The second is the Privacy Act of 1974 (USC 5 Sec. 552a) which was established to protect the rights of the individual. These two laws serve as a check and balance system on being able to provide data and information to the public while still preserving critical confidential data.

Oregon State Laws and Agency Policies

There are numerous state regulations in Oregon protecting the individual's confidentiality with respect to public records. Most of these relate specifically to the particular agency, its particular types of data, and its specific use of the data. Table 2 presents some of the major data themes that have confidentiality statutes directly impacting potential GIS data.

Table 2. GIS themes currently under confidentiality statues.

GIS Data Themes	Agencies	Statues
Endangered Species	ODF, ODF&W and USFS	ORS 496
Historical Sites	SHPO, ODOT and OR&RD	ORS 390
Pesticide Application	Agricultural	ORS 634
Employment and Workforce	Employment	ORS 657
Social Services	DHS	
Taxes/Income	Revenue	ORS 314
Workman's Compensation	DCDBS	ORS 697
Criminal Justice Data	Corrections	ORS 192

Case Studies

To review confidentiality and GIS data in Oregon, we have chosen two agencies for case studies. The confidentiality issues encountered here are likely similar to those encountered in other state and federal agencies.

Employment Department

The Oregon Employment Department (ED) collects and maintains data related to employment, unemployment insurance, and childcare. All of this data contains address-based locational information which can be geo-coded by street address matching.

There are several state statues relating to and controlling ED confidentiality. The primary one is ORS 657.655 "Confidentiality of information from employing unit records." This statute includes the following (bold type added only for emphasis):

(1) *Information secured from employing units…*

(a)*Shall be confidential and for the exclusive use and information of the Director of the Employment Department in the discharge of duties….*

(3) *Notwithstanding subsection (1) of this section, information secured from employing units pursuant to this chapter may be released to agencies of this state, and political subdivisions acting alone or in concert in city, county, metropolitan, regional or state planning to the extent necessary* ***to properly carry out governmental planning functions*** *performed under applicable law.* ***Information provided such agencies shall be confidential and shall not be released by such agencies in any manner that would be identifiable as to individuals, claimants, employees or employing units.*** *Costs of furnishing information pursuant to this subsection not prepared for the use of the Employment Department shall be borne by the parties requesting the information.*

Thus the statute allows the release of confidential data to certain entities, for planning purposes. Even then, the party that receives the data is bound by the OED public release of information rules.

The Research section of OED applies three rules to ensure that confidential information is not released. These are consistent with rules used by the U.S. Bureau of Labor Statistics, with whom OED works collaboratively on a number of statistical programs.

1) As required by statute, information on individual entities (persons, employers) is not released.

2) Aggregated information is released only:

   a) when there are at least three entities in a particular aggregation and

   b) when no one entity comprises more than 80 percent of the total.

Recently, an OED GIS workgroup reached consensus, based on the current statutory language, that individual (firm-level) geo-coded data cannot be shared with other entities (except for planning purposes) and that maps of point data (e.g. individual companies), at any scale, cannot be released publicly.

The primary rationale behind this decision is that while scale, time, individual knowledge, and other factors may impact an individual's ability to identify confidential information from a map, the existing statute simply states that "information received from employing units … shall be confidential".

It is extremely important, though, that data is available to policy makers and management in allocating resources. It is also important that governmental entities be as efficient as possible in their operations. Duplication of effort … for example, numerous agencies coding identical or similar data sets … is a poor use of resources.

Agriculture

Agricultural data is collected and maintained by the Oregon Department of Agriculture (ODA). Some of this data is publicly available and in GIS format. Confined Animal Feedlot Operations (CAFOs) are an example of an ODA GIS dataset that is commonly shared between agencies and organizations. This sharing of information means better decision making, resource allocation and reduced costs in not replicating efforts. However, some data, collected and maintained by ODA is confidential in nature.

Pesticide Reporting

The Pesticide Use Reporting Law requires all pesticide users (except households) to report their use of insecticides, herbicides, fungicides, rodenticides, and other pesticides to the Oregon Department of Agriculture (ODA). The information collected on each pesticide use report is as follows:

1. Date of pesticide use
2. Site Category and Specific Site (to indicate crop, or target area)
3. Whether the application took place on public or private property

4. The geographic location (including TRS, GPS, Address or ZIP Code)
5. Quantity and identity of pesticide products
6. Purpose for pesticide use

Pesticide users enter business and contact information and file use reports into the Pesticide Use Reporting System (PURS) database via a secure website interface. While under ODA control, the data is maintained such that only specific ODA employees and contractors who sign a confidentiality agreement may access it. The Pesticide Use Reporting System (PURS) was established under ORS 634 and OAR 603-057. ORS 634 is the State Pesticide Control Act. Section 634.042 deals with issues of confidentiality.

The pesticide use reporting law requires that ODA prepare an annual report of pesticide use summarized to state, county and 4th field HUC geographic areas. While doing so, ODA must keep information confidential that may identify the pesticide use of any one particular entity (except on applications made to public buildings, roads, or other public property).

The Law also allows for universities, state and federal agencies, and research institutions to request more specific information from PURS. In order to obtain this information, the interested party must submit a written request with specific information as detailed under Oregon Administrative Rule. The request must include an ODA-prepared confidentiality agreement and the entity's plan to protect the confidentiality of individual reporters. Breaching confidentiality of information from PURS could result in a civil penalty of up to $10,000.00.

Other Pesticide Related Activities

Nationally several states already summarize pesticide use and application data by county. Nationally the Census of Agriculture on a five year cycle and provides summary information on major agrichemicals applied with acreages and number of farms on a county level. Some states, such as New York, have already implemented statewide pesticide reporting programs. This is used as input to water quality modeling and evaluating environmental risks. Other states, such as California, have detailed downloadable breakouts by county and commodity type (application amounts, areas and chemical type).

Wisconsin is also in the process of developing a pesticide reporting system and has legislation pending in the current legislature. This will tentatively have a GIS Internet interface for visualizing where pesticides are applied, and the interface will be available to the public. More information on this is available at: http://www.wsn.org/pesticides/PDS_Components.shtml

The National Institute of Statistical Sciences (NISS) has also evaluated methodologies for agricultural statistics (information is available at www.niss,org/dg). NISS promotes using one of two rule based systems:

1. N-rule (minimum of 3 observation in any geographic or aggregation units)

2. P rule (minimum percent [p= 60 % of total acreage of all farms in unit])

NISS has noted that at a county level over 50% of the data statewide are not releasable and/or disclosable. These methods are similar to rules currently used in other state agencies such as the Oregon Employment Department.

Suggestions for Resolving Confidentiality Issues

Principle Should Drive Policy

1. Principle: Quality information allows informed decision-making.
2. Principle: Cooperation, coordination, and avoidance of duplication between governmental entities are worthy goals.
3. Principle: Individuals/businesses have a right to expect that their proprietary information will remain confidential and state agencies have a high commitment to meeting that expectation, according to all applicable legal requirements.
4. Principle: If state agencies have information that will be useful to public policy decision-makers, and if that information can be shared without causing damage to individuals/businesses' needs for confidentiality, then the information should be shared.
5. Principle: There is something of an inconsistency in allowing the sharing of confidential information for "planning" purposes but not for other "analysis" purposes.
6. Principle: If current statute places inconsistent or unreasonable limitations on our ability to appropriately share confidential information, we should seek changes to the statute.

Most, if not all, the statutes governing confidentiality have been in place for years. Do they still meet their intended purpose? Can they and should they be modified to allow more efficient use of shared information for policy and decision-making?

Possibilities for Consideration

There are numerous activities that can impact the future of how GIS and confidential data are treated in the State of Oregon. Some of these might be to:

- Form an inter-governmental group to examine confidentiality across all agencies, in the light of new technologies and methods for sharing and integrating data.
- Develop GIS Roles and Responsibilities related to confidentiality.
- Seek GIS data sharing and exchange agreements between agencies

- Investigate development of a uniform state government wide data sharing agreement.
- Develop written policy and guidelines for disclosure of GIS based data.
- Continue review and endorsement by
    - Geographic Project Leads
    - Policy Advisory Committee
    - Oregon Geographic Information Council
    - Attorney General
- Consider changes in current statutes to allow for the sharing of confidential information:
    i) Consider changes to individual agency's confidentiality statutes. e.g. Modify the Employment Department statute to allow sharing of confidential information not just for "planning purposes", but also for GIS-related purposes.
    ii) Consider proposing a specific GIS confidentiality statute that would, with appropriate safeguards, take precedence over any other state statutes or administrative rules.

Oregon Geographic Framework Implementation Teams

Established by inter-agency agreements between several State, Federal, local and private organizations, Oregon's Geographic Framework Implementation Teams promote the coordinated development, use, sharing, and dissemination of geographic data.

In the fall of 2000, a diverse group of individuals and organizations from the Oregon Geographic Information System (GIS) Community participated in a forum on GIS issues in Salem, Oregon. This effort initially defined the major framework teams and their elements in Oregon.

For more information about the Teams, or to be added to the framework mailing list, please contact:

Oregon FITS

c/o Oregon Geospatial Data Clearinghouse

955 Center St. NE, Room 470
Salem, Oregon 97301-2556

Telephone: (503) 378-6066

Facsimile: (503) 378-5200

Internet (electronic mail): cy.smith@state.or.us

World Wide Web: http://www.sscgis.state.or.us