

Task 1:
Evaluation of Data Management Systems for
San Diego Management and Monitoring Program Data

INTRODUCTION

The San Diego Management and Monitoring Program (SDMMP) was recently created to coordinate reserve management as well as species and habitat monitoring efforts in San Diego County. Currently such activities are undertaken by several local, state and federal agencies along with non-profit organizations and consulting firms. Effective management of the data coming from the various sources will require a robust data management system that centralizes storage, viewing and analysis of the data and effectively addresses the various conservation issues facing San Diego County.

The task of managing biological data is typically undertaken on a project by project basis. Project-based data management can lead to difficulty in developing consistent datasets when conducting large-scale trend analyses and typically does not promote data sharing amongst research groups. San Diego County-wide monitoring and habitat management will likely require regional trend analyses as well as access to data collected from various sources. Several agencies in California have developed data management systems to combat these issues and provide a one stop shop for biological monitoring data. Within San Diego County, three examples of these data management systems include California Department of Fish and Game's (CDFG) Biological Information and Observation System (BIOS), U. S. Geological Survey's Multi-taxa Database (MTX), and the County of San Diego's SanBIOS.

Two priority issues that the SDMMP has been tasked to work on are MSCP Connectivity Monitoring and Land Management data. Several types of studies have been proposed over the next few years including carnivore tracking, analysis of genetic data, and bird banding studies. The three data management systems were evaluated on how each would manage these types of data and determine which, if any of the systems, would be best suited to house data collected for and by the SDMMP.

BACKGROUND ON EXISTING DATA MANAGEMENT SYSTEMS

BIOS

The BIOS data management system is run by the Biogeographic Data Branch (BDB) of the CDFG, based in Sacramento, CA. BIOS is similar to the California Natural Diversity Database (CNDDDB) except that BIOS incorporates data on all species in a non-standardized format and each dataset is its own layer in the viewer whereas CNDDDB only presents standardized rare species data. BIOS also presents associated environmental data if submitted along with the biological observations.

The following description of the system is taken from the CDFG BDB website (CDFG 2010):

BIOS is a system that enables the visualization of the spatial distribution of biological data generated by the Department of Fish and Game (DFG) and its Partner Organizations, the management of those data when necessary, and the sharing of those data with Department employees and partners. BIOS uses standard guidelines, protocols, and tools that enable the analysis and management of field observation data. Notable features include:

- Recommendations on how to build, collect and store datasets that can be used together,
- Data warehouse to receive datasets, store them in a consistent manner, and serve spatial and attribute data associated with biological observations,
- Catalog and tools that enable online queries based on attributes and spatial location,
- Management system to create and maintain GIS features,
- ArcIMS (Internet Map Server) to view, query, and retrieve biological and spatial data from BIOS, online. [Currently, CDFG is in the process of moving BIOS to ArcGIS Server.]

The BIOS map viewer presents spatially enabled data via the CDFG BIOS website. The data can be viewed along with other selected datasets. Maps can be printed off of the website for inclusion in reports. Currently over 400 datasets are on BIOS and include vegetation mapping, proposed and final critical habitat maps from USFWS, tree delineation, inland fishery data, invasive species data, etc. There are minimal requirements for submissions. Primarily, submitted datasets must contain bio-spatial data and adequate metadata must be included, detailing the specifics of the dataset as well as Contributor contact information. There are currently 2 levels of access, a public and private viewer (for sensitive species). The private viewer requires a Login and password, which is only given to CDFG and associated partners staff. The Contributor must indicate the accessibility level (public, government-only) and download access of each submission (via BIOS or providing specific contact information). If the contributor gives permission for data to be downloadable, it will be in a CSV or DBF format; the original shapefile cannot be obtained from BIOS, only from the contributor. The viewer in itself can be used to conduct spatial analyses of species presence or habitat data. Additional analyses may depend upon the format the data was submitted in and could require working directly with the contributor.

BIOS is essentially a data warehouse, it presents a collection of datasets which can be queried individually but does not allow for queries across multiple datasets. Data submitted to BIOS is still managed by the contributor and typically persons interested in a BIOS data layer must contact the contributor for questions or data requests. To submit data, a shapefile or geodatabase is preferred. There is typically a backlog so posting of data submissions can be several months unless one specifically requests that the data be posted sooner.

While BIOS offers several benefits as a data management system, it does have some drawbacks. If this option is used as the sole data management system, each entity that conducts management and monitoring efforts would still house and manage their own data in most cases. This could potentially make it more time consuming and difficult in compiling and standardizing the multiple datasets into a single analyzable dataset.

MTX

The USGS Western Ecological Research Center's San Diego Field Office EcoInformatics Program has been developing a data management system, currently called the Multi-Taxa Database (MTX). Development of this regionally integrated database began in 2002 in partnership with California Department of Fish and Game with legacy data migration beginning in 2006 continuing in partnership with U.S. Fish & Wildlife Service and National Biological Information Infrastructure (NBII). This integrated relational database allows for storage and retrieval of protocol driven ecological data using many standardized and customizable tools. Data can be collected, entered, reviewed or reduced using PDAs, desktop and laptop computers with common software (Webforms, Access/Excel, ESRI, etc.) (Holmes and Brown 2010).

The MTX currently houses data on Herpetofauna and Ant Pitfall Trappings efforts, Carnivore Camera and Scent station monitoring data, Vegetation Point sampling, Satellite Telemetry tracking of birds and Pocket Mouse survey data. Associated habitat, weather, observer, water chemistry data can be stored as well as many other aspects of the biological sampling data collection. As USGS migrates additional datasets into the MTX, staff continually update the database structure to accommodate the new datasets. The database can be linked to ESRI's ArcGIS software so spatial analysis and QA is easily conducted through ArcGIS. The data can also be queried and manipulated in a format for use in statistical analysis; the queries can then be exported in a file format readable by other software.

The database is stored in SQLServer and is managed and maintained by staff located in San Diego, Dixon, and Sacramento, with the actual database server located in Sacramento. Data can be submitted for migration as an Excel spreadsheet or Access database. Capability exists to load data directly from a PDA, but specific PDA forms need to be developed based on the project data beforehand. Currently the MTX database is viewable to USGS and other DOI agency personnel. Issues with the USGS firewall and security policies need to be addressed in order to make the viewer accessible to other interested parties or the public. Access to the database from other DOI offices has been tested and works, although the connection is often times slow. The MTX data management staff are aware of the issues and are intending to move toward use of WebForms for data viewing, which should resolve the connection issues. In addition, MTX staff are currently developing user access roles that will enhance database security. Users will be given task-specific permissions based upon their role within the project (data entry, QA, or management). The database provides many benefits as a data management system, with most negative issues related to security and staff time issues. If used as the data management system for the SDMMMP, any type of study would be easily uploaded into the database, adequate viewers can be easily built for the data, and analysis of the data would be accomplished by building queries. The issues of access due to firewall security constraints may be an initial problem but the USGS staff are currently investigating ways to deal with this issue.

SanBIOS

The County of San Diego Land Use and Environment Group has been working on a data management system to manage the data they collect and receive. They received guidance on data management systems from the USFWS in 2006 and created SanBIOS, an ESRI geodatabase, which was designed to be compatible with BIOS. The following description was taken from the County's SanBIOS User Manual (County of San Diego 2009):

Created in 2009, the SanBIOS database serves as a single repository of species observations collected by various departments within the County of San Diego's Land Use and Environment Group. Coordination of biological species tracking between departments ensures a complete dataset, meeting a specific data standard, and will provide the best available information to environmental scientists, advocacy groups, all County departments, and various agencies.

This catalog of species observations have been recorded by professional biologists from the County of San Diego as well as from various other agencies and private firms. These data serve many functions: they serve as a baseline catalog of species records in the adopted South County MSCP and the draft North and East County preserve systems (including invasive species) in the incorporated and unincorporated areas of San Diego County, they are used to direct the location of future permanent plot surveys and for uses in various monitoring projects, they are used in the testing and validation of predictive species niche models and to comply with CEQA regulations, and they are used to identify and monitor invasive species. It is important to note these observations are an indication of confirmed species presence at the time of the survey but offer no indication of species absence. The types of surveys performed to collect these data were variable, ranging from highly organized and standardized surveys to random observations based on chance. SanBIOS was constructed to be interchangeable with the State of California's Biological Information and Observation System (BIOS) database (<http://bios.dfg.ca.gov/>).

SanBIOS is available for download through SanGIS (<http://www.sangis.org/>) and is physically stored in the County's Data Warehouse. Currently the geodatabase has approximately 80,000 records and is 26 MB in size. An interested party, who can be anyone from the public, must first request an account to access SanGIS data and then can download the entire SanBIOS. While metadata exists that describes what each field in SanBIOS means, no metadata exists that describes the data that comprises the geodatabase. Spatial analysis of the SanBIOS data can be done through ESRI ArcGIS software. Persons without this software may have to work with the County staff to get the data in a format useable to them. In addition, the data source/contributor may need to be contacted regarding the metadata before using specific data in any analyses.

Data contributors can download a geodatabase template to use as guidance when submitting data. The standards used to develop the template were based upon the data needs of various County departments that collect GIS data. Each field in the geodatabase must contain data or the designated NULL value. The County will no longer accept data that does not conform to their SanBIOS standard. Additional data can be collected for projects but will not be stored as part of the SanBIOS geodatabase. As stated earlier, the public has access to this data; it is the contributor's responsibility to ensure that sensitive data is treated appropriately before being made available to the public. The SanBIOS staff will QA the data to the extent that it conforms to their geodatabase standard, otherwise the QA is the responsibility of the Contributor before submission to the County. As with the previous data management systems, SanBIOS has both positive and negative aspects as a data management system. If SanBIOS was used as the sole data management system, stakeholders are limited in what data fields are viewable and

downloadable in the geodatabase. The SanBIOS staff do not update the data from outside sources other than what is submitted to them so certain datasets within SanBIOS may not up to date (for example, CNDDDB). In addition, the lack of metadata on what datasets comprise SanBIOS could cause issues if using the data to conduct large-scale analyses.

SDMMP DATA MANAGEMENT RECOMMENDATIONS

The SDMMP has indicated that any database they use needs to encompass more than species occurrence data and must meet, at minimum, the reporting requirements for SanBIOS and the USFWS Carlsbad Fish and Wildlife Office (CFWO). They also expressed that the database must meet the needs of stakeholders such as land managers and biologists who will be involved in the connectivity monitoring and other future studies.

The SDMMP has proposed various methodologies to accomplish the MSCP connectivity monitoring tasks, such as GPS telemetry studies, camera and track station monitoring, bird banding and analysis of genetic samples. Land management data that the SDMMP is currently evaluating for database management includes population monitoring and land restoration actions. These studies include tracking of species presence and absence from a reserve, removal of invasive species and replanting of native species.

Currently, BIOS and the MTX house datasets similar to those proposed for connectivity monitoring and could effectively house these data with little or no ‘data massaging’. Both systems currently contain carnivore camera station and tracking data. Thus, it seems logical for the SDMMP to adopt these systems for their data collection as opposed to creation of a new database. Further, the MTX has tables and fields already developed for genetic sample collection and bird band identification data in addition to presence and absence data for species tracking. While the MTX does not currently house much land management data, a LandManagement table has been established and can easily be updated with additional data as it is introduced for import into the database.

Creation of a new database system requires considerable time and research in development of a schema and management protocols. Since data management systems that can effectively handle the potential SDMMP data needs are already in use in San Diego County, there is no need to “reinvent the wheel” and develop a stand-alone database. The SDMMP is still developing and finalizing other types of studies where data will be collected and need to be managed, and development of a new large-scale database system would not be feasible until all protocols and potential data fields are known. To satisfy their immediate data management needs, they should choose an existing database and follow the standards established for that system.

Logistically the adoption of the USGS MTX database is the best option for the SDMMP for several reasons:

1. The MTX database already contains the necessary tables and fields for many biological monitoring protocols including those proposed for Connectivity Monitoring. In addition, some

Land Management tracking data can be loaded into the database. It is relatively simple to update the tables to accommodate additional data that is not already included in the MTX.

2. The MTX TaxaObservation table has a SpeciesDetected field, which is useful for projects that target specific species. This allows for the storage of negative sightings data.
3. USGS and SDMMMP share the same office. This is ideal for optimal communication between entities as well as resolves the USGS firewall issues for the SDMMMP staff.
4. The MTX has minimal data standards thus the SDMMMP are not limited by what data they intend to store in the database nor are they required to include data that is not pertinent to their interests.
5. The MTX offers the potential to enter data in multiple formats, via PDA, or MS Excel or Access uploads. USGS is currently investigating use of WebForms for data entry and viewing.

Using BIOS or SanBIOS would be insufficient as the data still needs to be in a centralized location and must be standardized to perform complex analyses beyond spatial analysis of occurrence points. If the SDMMMP adopts the MTX, BIOS and SanBIOS should still be utilized as both systems are beneficial in disseminating data to interested parties, as the USGS MTX is not currently available to the public due to firewall issues. The BIOS team could potentially provide SD County with an online viewer specific to the local interest (akin to CalFish, <http://www.calfish.org/>). This may be a better option than SanBIOS as BIOS presents the data as submitted by the Contributor as opposed to a mandated format.

The CalFish program has a website detailing the project and partnerships. They provide interested parties with downloads of data and shapefiles accessible through a series of queries. If a visitor selects the option to “View maps”, they are navigated to the BIOS viewer with the CalFish datasets available for selection. The SDMMMP website (www.sdmmmp.com) could include a weblink to BIOS much like CalFish where the visitor is not even aware that they are accessing the CDFG website.

REFERENCES

California Department of Fish and Game. Biological Information and Observation System. <http://bios.dfg.ca.gov/>. Accessed 7/29/2010.

County of San Diego. 2009. County of San Diego SanBIOS GIS Data Standard User Manual.

Holmes and Brown. 2010. Data Development Projects for WERC EcoInformatics. Information Sheet. Prepared April 2010.