California Department of Fish and Wildlife
Field Data Collection Database Template Introduction

Prepared for the Field Data Collection Work, 11/12/2007
Last updated by Douglas Burch, 12/3/2014

Overview

The purpose of this document is to provide some basic information about the Field Data Collection Database Template (template) concept and explain how it works. The template is the synthesis of various efforts to standardize the way that field data is collected, entered, and managed within the California Department of Fish and Wildlife (CDFW) at the field level. In particular, this database model focuses on species centric field data, such as that collected by the CDFW during its various monitoring efforts, but it could be extended to other categories of data as well. The main goals of the effort to develop a standardized field data collection model are to help provide a departmental standard for field data collection and to bridge the gap between highly centralized databases and those of distributed databases that are better suited to smaller business model projects where data is collected in the field and typically must be managed for a time by the project lead.

It may first be necessary to address why a departmental standard is desirable. In some cases a standard is not necessary, and the intention of this document is not to recommend the template as a solution in every case, however in most cases having a template to use at the onset of a data collection effort has many benefits. For projects where data sharing and merging of discrete datasets is inevitable, adhering to a departmental standard is critical for both data integrity and comparison – not having a standard will likely result in incompatible datasets even though the various data collection efforts address nearly identical objectives thus significantly limiting how the data can be used. In addition, a template will save time by facilitating implementation of the various steps involved in setting up a field data collection program, for example, planning, identification of data needs, data collection, data entry, data compilation, data management, data delivery, data use, and data analysis. It is one of the goals of this effort to not only provide a template for facilitation of these various steps, but it is also for the intent of facilitating a collaborative process for sharing ideas and wisdom within the CDFW as well as among collaborating agencies, that is, combining efforts between the greater field data collection and management community.

An inherent goal behind use of the template is to move data from its point of origin – typically a field survey data collection scenario – to a central repository, such as a secured web-enabled data server. The additional goal of not encumbering field data collectors with complicated database architecture – while forcing data integrity necessary for eventual inclusion of field data into a central, secured data repository – is a challenge. It may not be the solution in every case, but the template does provide a workable approach for bridging the gap between needing simplicity in the field data collection scenario and scalability, that is, the ability to normalize raw field data so that it can be stored more efficiently in an all-inclusive datastore. Perhaps more

than anything else the template is a process for communicating ideas – comparing knowledge and experience with others – about how we approach the field data collection process and management of the data which the process generates. In other words, we are attempting to avoid having multiple field data collection efforts that are continually "reinventing the wheel." Since we collectively benefit from individual experiences and knowledge, it stands to reason that products derived from field data collection activities which make use of the template can be something that is more than the sum of its parts.

Template Basics

The template structure by intention is generic enough so as to build-in flexibility, and it uses what has been coined in earlier conceptual efforts as an "observation-based" approach. This approach or process defines field data collection – particularly where there is a species or organism focus – in terms of "units of observation," originally outlined with working examples in the document titled *Uniform Field Observation Model -- Specifications for Minimum Uniformity Requirements*, an internal departmental document (Burch, D., Gaul, P. A., Haney, E., Davis, K., and Kauffman. 2001), and a synopsis of that is available in the document titled *Observation-based Approach to Developing Databases* (Burch et. al., 2007). To some degree units of observation are just organizational rules intended to aid in managing large, seemingly disparate, collections of data; however, there is a practical and rational basis to this organizational construct, and units of observation can agree well with the standardized, and usually routine, procedures used by the CDFW when collecting field data, particularly for monitoring efforts where there is a biological concern, for example, in a fisheries study a unit of observation might be an event like the "pulling of a fish trap" out of the water, and information about the trap itself would be a data object that is classified as the "observation event" while what is observed in the trap are the details of that observation event, or its sub-data objects or components.

In using the template we define an observation event as "Header[1]" information, that is, information which precedes a sub-data component and is the same for all of its sub-data entities or records. With each observation event, typically there is detailed information – about the individual organism – in a one-to-many relationship, for example, "Individual" data might include measurements related to an individual organism while in cases where any number of individuals have been counted, and "batched" together as a single record for quicker processing, there would be "Count" data. In addition to all of the attributes associated with an observation event, or Header record as it is defined in the template data structure, it is also fundamental to the template as a model that we maintain both standardized location information about where the event occurred and project level metadata to describe any "Project" specific protocols. The three of these generic data components, that is, (1) unit of observation, (2) standardized location, and (3) project level metadata, are core requirements of

[1] For the purpose of this document we use capitalization to indicate that words, such as "Header," "Individual," or "Count" are being used as proper pronouns, for example, in this case "Individual" represents a data object class.

the observation-based approach which can be implemented when using the template as a strategy for organizing information collected while conducting field survey work for any given area of environmental study. We can also add a "Survey" level to this hierarchical organizational structure, that is, we can group all units of observation that share the same spatial extent, same time period, and are part of the same project or scope of effort. Since these "core" aspects of the template structure have been established and standardized, any number of sub-data classes (tables) and properties (columns) can be added to this foundation, for example, in the case of our example above a "tag" table could be added as a "child" table to the Individual table, its "parent," and "tag-identifier," "tag-type," and "tag-color" columns could be added to capture properties or attributes of each tag observed on or in an individual.

While this "add-on" capacity of the template provides for a flexible quality to the data model, the foundation or "core" aspects adheres to a strict standard and is not subject to modifications or customization – in other words, the core structure does not change, allowing for trend analysis quality data that can be useful across a broad landscape (Figure 1). In addition to the tables mentioned in the earlier example, there are quite a few more detail level tables which maintain attributes specific to different types of surveys, and there are a sizable number of reference – or lookup – tables which support data in the template structure. Attributes of the core structure capture generic information in a standardized format, that is, the "who," "what," "when," and "where" of the observation event, but in addition to the core structure there is what has been coined the "cafeteria" tables. These are tables that maintain all the attributes or data elements that a field data collection effort could possibly need or want to collect, and these data elements can be added to a working database structure as needed for specific project needs. If the data element does not already exist within one of the template's cafeteria tables then it is added and becomes a part of the template; this new data element then becomes available to current and future field data collection efforts implementing the template as part of a field data management strategy.

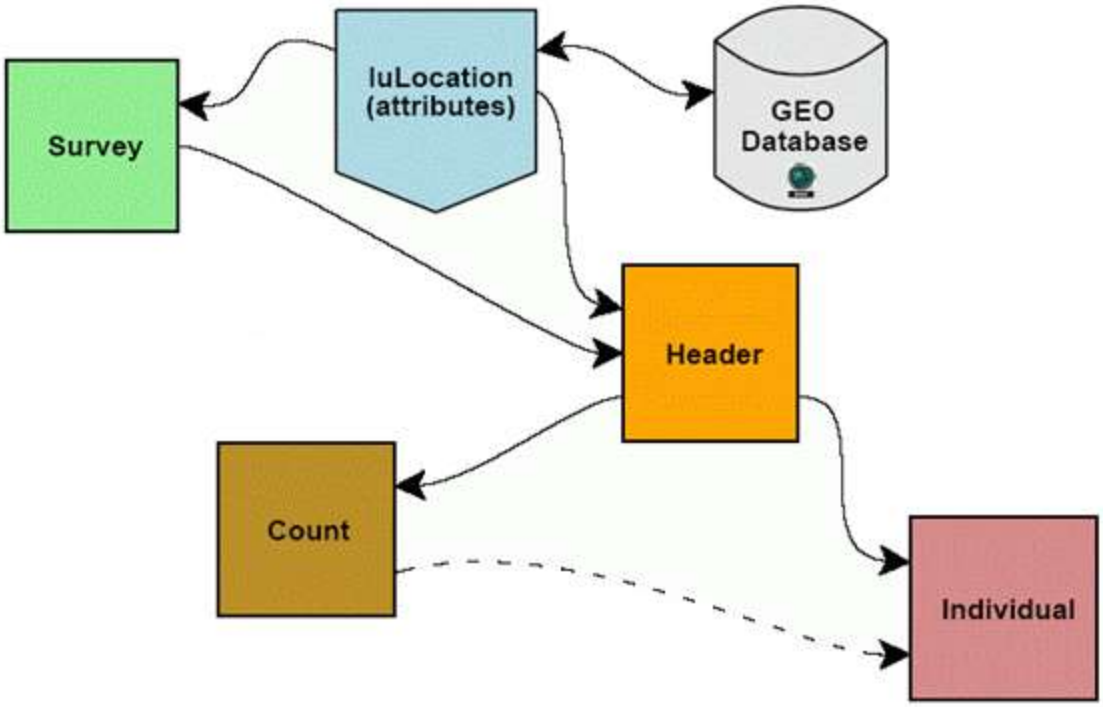

Figure 1. The diagram is a simplified illustration of some of the Field Data Collection Template's core tables along with their relationships to each other. Solid lines indicate relationships enforced by the database design while the dashed line implies that the linkage is conditional and enforced by programmatic data rules within the database's frontend application (its user interface tier).

Taking a step back and looking at the template design at a higher level, the organizational hierarchy has three more nodes or relationship levels above the Survey table. These high level nodes in the structure allow for maintaining data from multiple projects and programs in a single database.

Illustrated in Figure 2, the "Observation" node of the template model is equivalent to the Header table, and the "Details" node represents its sub-data tables. The Project level consists of a simple lookup table which is a list of possible projects that represents purpose or scope of the field data collection effort; all data must be a part of an established project within an overarching statewide program, for example, coastal versus inland monitoring programs. Since this is for ultimate inclusion into a statewide database then for the purpose of the template, the State is the highest level of the organizational hierarchy. A couple other aspects of the diagram worth pointing out are that (1) nodes above the Survey node are based on business management entities, such as a management unit assigned to a specific project lead and defined spatially by something like a district or management boundary. And, (2) there will be only one record in the Metadata table per project, however if methodology has been adopted according to programmatic protocols then there should be little difference between the metadata of one project and another within the same Program node.

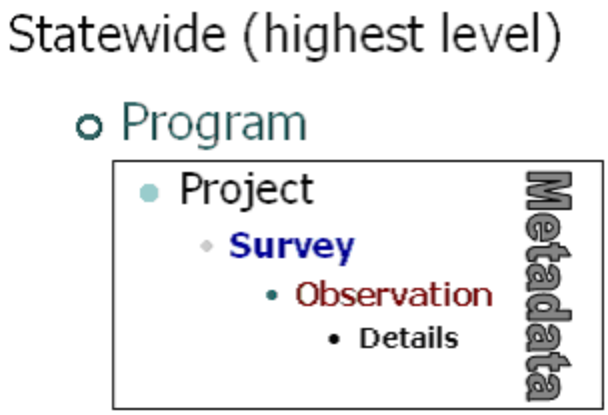

Figure 2. The graphic illustrates positions of the various hierarchical nodes within the Field Data Collection Database Template's organizational structure.

There is also a "layered" quality to the template's hierarchical levels of organization, its database components, and the relationships between those components (Figure 3).

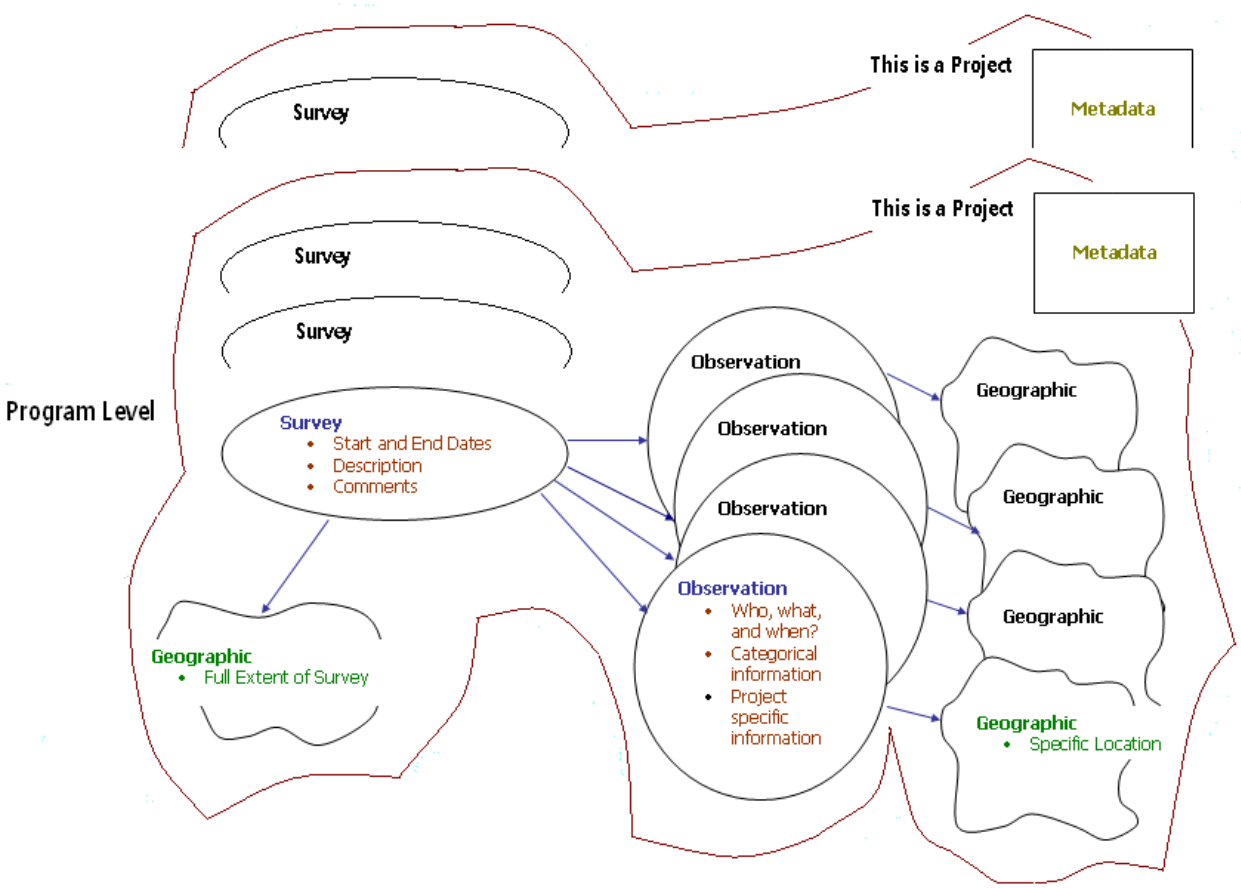

Figure 3. The diagram is a high level representation of how data is organized and "layered" by the Field Data Collection Template structure. There can also be multiple entities at the program level, such as an "inland" program versus a "coastal" program (not shown here).

A benefit from the observation-based approach, implemented through use of the template, is its ability to organize field data in the form of discrete packages (Figure 4) to better manage the data as subsets of a larger effort yet the subset can be a complete set or package of information itself. This model also allows for the data to be highly portable.

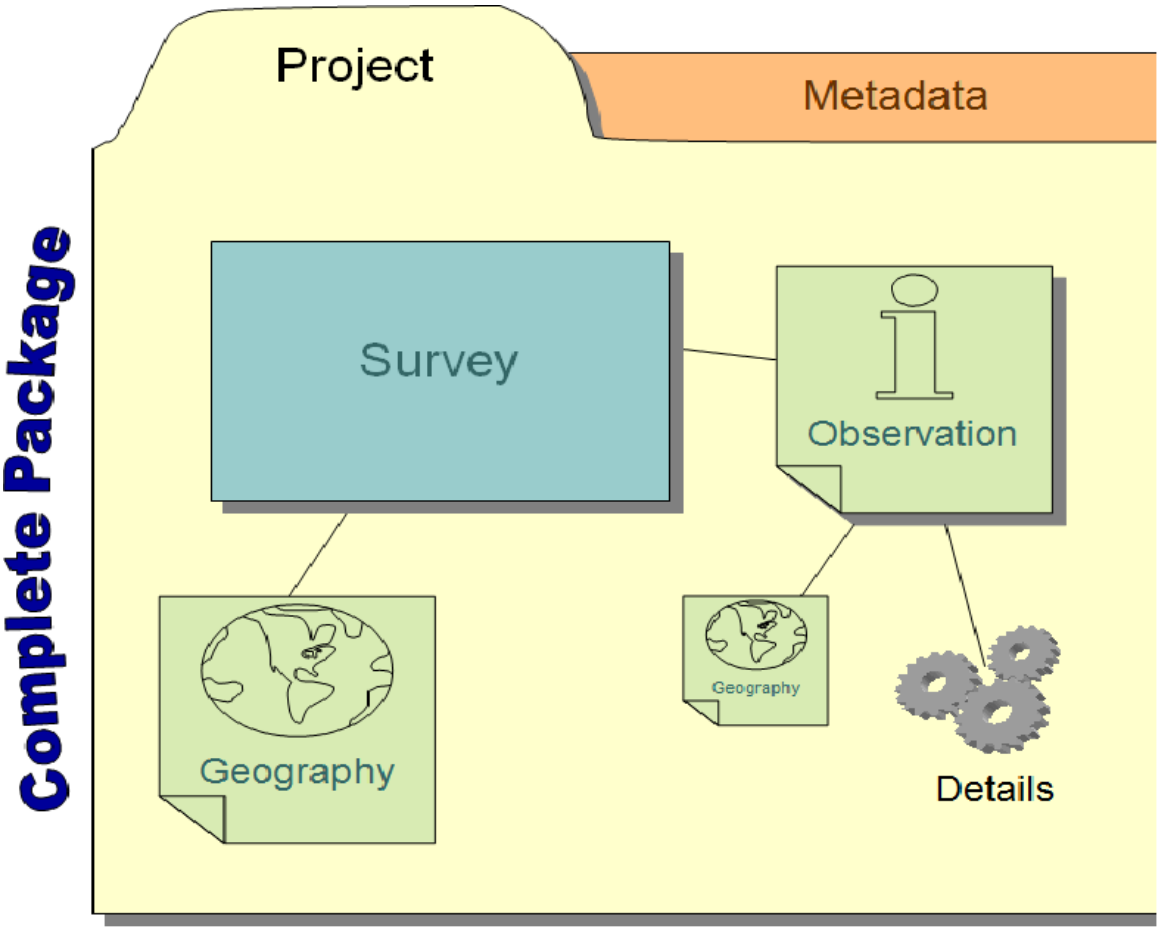

Figure 4. This figure illustrates how the Field Data Collection Database Template strategy allows for managing large amounts of incoming field data as a series of data "packages." These data packages can be managed locally if need be while surveys are underway to eventually be merged with other packaged data in order for a centralized program level dataset.

Using the Template

The template is a concept as much as it is an actual data structure. It has been fully documented in the form of a data dictionary and an entity diagram that is capable of generating documentation about itself; it is also available as a portable database structure that can be used as is, or, more likely, simplified to work in a particular field data collection scenario, for example, Individual and Count tables can be combined for use in a scaled down hand held data collection application requiring a simplified or more flattened database design. The main caveats behind using and customizing a database structure based on the template is to always retain field names, field data-types, and field definitions exactly as they are in the latest

template version so that the field data structure can be scaled up to more efficient normalized formats when it becomes time to merge and manage the data in a centralized datastore.

In closing, the template to a large degree has been modeled after existing enterprise level natural resource database architecture as this would likely be the ultimate repository for warehousing centrally managed State natural resource data. That being said, the template development has been driven by the needs of those collecting data while keeping in mind a focus on the ultimate need to have all of this data in one secure place.

Citations

Burch, D. Observation-based Approach to Developing Databases. California Department of Fish and Game; 11/12/2007. [Cited 2014 December 16]. Available from: https://nrm.dfg.ca.gov/FileHandler.ashx?DocumentID=92314

Burch, D., Gaul, P. A., Haney, E., Davis, K., and Kauffman, E. Uniform Field Observation Model Level I: Specifications for minimum uniformity requirements - version 1.0c. California Department of Fish and Game; 11/1/2002. [Cited 2014 December 8]. Available from: https://nrm.dfg.ca.gov/FileHandler.ashx?DocumentID=91757