Pacific States Marine Fisheries Commission
People-Project-Organization Schema: PPO
CONTRACT NO. 06-71 - TASK ORDER NO. 06-08

Doug Clough, SYNERGETICS Engineered Systems
Ver 1.1, 10 Oct 2006

INTRODUCTION

Task Order 06-09 budgeted 178 hours for development and implementation of a data model representing:

- organizations, research projects, and their relationships with PTAGIS
- people and their roles within these organizations and projects
- PTAGIS data-sets and their custodial owners

The model must be capable of tracking these relationships through time, as people move between organizations and roles are re-assigned.

To fulfill these objectives it's essential that we define precisely what constitutes a **project**, and the **roles** fulfilled by its participants. In particular, we must equip our model with the means to characterize each of the possible data-set types, in sufficient detail to answer the following questions:

- Given the name of a person who is custodial owner of a data-set, what attributes define the body of data for which they are responsible? Where do the associated data reside?
- Given a data element – e.g. an interrogation record, or a tagging record – who is the current custodial owner? What is the history of custodial ownership?

Version 1.0 of this document differs from its predecessor, Version 0.4, by incorporating changes proposed in the PPO 'wiki' Ver 0.4 Reader Comments, on modeling organizational changes, and on dealing with interrogation super-sites. Version 1.0 is the first implementation candidate for the PPO schema. As described in the PPO Road Map, the logical ER diagram presented in this document will serve as the basis for the initial physical implementation.

"Green entities" of the PPO logical ERD will be implemented as database tables, and populated with data from existing disparate repositories. Possible development of a PPO web-interface, capable of performing Create-Read-Update-Delete (CRUD) operations on the associated database content, will be deferred until we've gained confidence in the schema, from the standpoint of operational practicality.

DEFINITIONS

Nouns essential to the Data Model discussed in this document are defined in Table 1, below. Notes accompanying a definition provide examples or warn of inconsistencies with legacy implementations of similar concepts.

Table 1 - Definition of Terms

Noun	Definition	Notes
organization	A state, federal, or tribal entity, or agency thereof conducting one or more PTAGIS **projects** during a stated span of time.	A person can be affiliated with more than one organization at a time. A person may move from one organization to another. The legacy PTAGIS system identifies organizations through an entry in **valid_tbl** (select * from **valid_tbl** where domain='ORGANIZATION';). The six character code is used in the ***_hdr** records defined in the PTAGIS Specification document. The TDI subsystem ignores the ORGANIZATION domain represented in **valid_tbl** and allows fifty characters for an organization identifier. This data inconsistency should be repaired during the implementation phase of this project.
role	A clearly-defined collection of responsibilities to be discharged by an individual during the span of time in which an **organization** conducts a PTAGIS **project**.	Roles are: Project Sponsor, Tag Coordinator, Site Coordinator, PTSC, ... In the legacy PTAGIS system, coordinators were identified through an entry in **valid_tbl** (select * from valid_tbl where domain='TAG COORD'). The coordinator code is three characters and is present in most of the ***_hdr** tables. Tag Coordinators are recruited by PTSC representatives and are responsible for collecting tagging, release and recapture information and submitting it to PTAGIS. Site Coordinators are recruited by PTAGIS and are responsible for data collection and equipment operations at a passive interrogation site. More recently, the **coord_id_info** table has been used (in addition to **valid_tbl**) to keep track of Tag Coordinators. An ancillary table called **coord_status** has knowledge of status indicators assigned to any of the **coord_id_info** persons. The **tdi_sponsor_coordinators** table was an attempt to model the hierarchical relationship between NPCC FWP Project Sponsors and Tag Coordinators. This is a convenient way to know which project a tag_coordinator is working on. I question whether this is useful, however, since one can link a tag_id from a tagging file to the project associated with the PIT Tag Distribution to get that information more accurately.
person	An individual affiliated with an **organization** charged with exercising one or more **roles** during a stated span of time.	The generalization “people” refers to a collection of **person** instances. Currently, there are various people collections in PTAGIS. For example, the PTAGIS LDAP repository contains people that have a PTAGIS user account. The **tdi_contact** table contains people that order or receive shipments of PIT tags. The **pab_contact** table contained a list of PTAGIS contacts of all types (this table, and all pab_* tables, are obsolete and should be dropped during the implementation of this project). The **tdi_project** table contains copies of contact information that overlaps the **tdi_contact** table. The **tdi_project** table should be altered to remove the contact specific attributes and replace them with a reference or key into the **person** table during the implementation of this project. A **person** should

		have a first, middle and last name.
activity	Generally, rearing, tagging, release, detection, separation-by-code, and recapture of salmonids (or other species), at *locations* and *times* detailed by entries in the ALRS sub-schema of the PTAGIS3 database	It will be useful to include **activity** and **activity_location** (from the ALRS sub-schema) in the PPO logical ERD, as landmarks at the periphery of the new conceptual landscape. Which of the possible relationships between PPO and ALRS tables we'll choose to exploit – if any – will be determined later.
project	A collection of PTAGIS **activities** undertaken by an **organization** at specified locations during a stated span of time, resulting in the creation of one or more PTAGIS **data sets**.	PTAGIS projects primarily include Northwest Power and Conservation Council's Fish and Wildlife Program projects funded by Bonneville Power Administration. Other projects may be incorporated secondarily. A PTAGIS project could be 'owned' or 'sponsored' by more than one organization. Version 1.0 of the ERD supports multiple organizations exercising a variety of roles with respect to a project.
data set	**A** - A body of information generated by a **project** as it conducts a PTAGIS **activity** at one or more locations during a stated span of time **B** - A body of information generated by a **project** as it utilizes a PTAGIS **activity_ location** during a stated span of time	As contemplated here, a data set could consist of tagging, release, interrogation, recapture, mortality, or tag-distribution data. Note that the **tdi** application associates a **person** with a 'distribution'. But how do we distinguish one data set from another? For example, do all of Doug Marsh's (DMM) tagging files constitute one data set? Have all DMM files been produced under a single project? Are data-sets 'location specific'? An interrogation record is highly location specific – with precision on the order of antenna width. How location-specific is an 'interrogation data-set'? How about a 'tagging data-set'? Definition **A** is non-location specific; **B** is location-specific. Which is it? In contrast with Version 0.3, which used a hybrid of the two definitions, Versions 0.4 and 1.0 of the ERD implement Definition **B**.
custodial ownership	A type of **role**, connoting responsibility for collection and maintenance of a PTAGIS **data set**	Custodial ownership of a data set is never vested in more than one person at a given point in time. We could define 'Tagging Coordinator' as a **role** that endows a **person** with **custodial ownership** of a 'tagging' **data set**.
custodial owner	The individual named in a relationship assigning **custodial ownership** of a given PTAGIS **data set** to a specific **person** during a stated span of time	If the custodial owner of a data set were to move from one organization to another, a new custodial owner would be assigned from within the original organization. PTOC is the 'interrogation coordinator' for main-stem sites, with 'custodial ownership' of interrogation data assigned to a member of the PTAGIS staff. For tagging, mortality, recapture, or small-stream interrogation data, custodial ownership resides with <activity> Coordinators of the submitting project. Note that the term 'Site Coordinator' would be synonymous with 'interrogation coordinator' under this formulation of the logical data model.

LOGICAL DATA MODEL

Based on the concepts discussed above, Figure 1 presents a fifth attempt at the PPO Logical ERD. Note that, by definition, a 'logical' model is studiously devoid of implementation details. True enough, it may help visualization during the design phase to imagine that each entity-type (aka "entity") in the logical ERD corresponds to a database table. However, this is not to say that every entity shown in the diagram will be implemented as a table – or, for that matter, implemented at all.

With that in mind, it can be useful to include some entities at the periphery just to show how they relate to the ERD's central ideas. When this is done, a closed boundary can be drawn on the ERD to distinguish clearly those entities ("inside") that are to be implemented from those ("outside") that are not. Often it's sufficient to capture just the names, coded identifiers, or "natural keys" of such "outside" entities, for use as attribute values when instantiating the "inside" entities with which they are conceptually associated.

For example, the existing **valid_tbl** incorporates an ORGANIZATION domain with a set of records corresponding to organizations that submit field-data files. The concept of **organization** is "outside" the implementation. However, the corresponding **organization codes** are captured for use with tables that record a minimal logical relationship, in which it's sufficient to store just an organization identifier.

From this perspective, the current effort can be seen as expanding the conceptual boundary of the original PTAGIS implementation. Originally outside the boundary, **organization** is now inside the boundary. At implementation time, we might choose to create an **organization** table and populate it with one row for each **valid_tbl** entry in the ORGANIZATION domain, at the same time entering values for all of the **organization** attributes. Once the disembodied organization codes in **valid_tbl** had been replaced by fully attributed records in an **organization** table, the ORGANIZATION entries would be removed from **valid_tbl**.

Similarly, the expanded boundary encompasses the notions of **person_role** and **organization_role**. Although in some contexts a role might embody numerous attributes, for our purposes it may be entirely sufficient to capture just two: a role-identifier – e.g. role name – and a description. If this were the case, we could implement **person_role** and **organization_role** simply by making two entries in the (already implemented) **ptagis_domain** table to define 'person_role' and 'organization_role' as domains, plus one entry in **domain_enumeration** for every role we want to define. Note that **ptagis_domain** and **domain_enumeration** accomplish the same thing as **valid_tbl** – but **ptagis_domain** lets you name and describe an enumerated domain independently of its value list.

Before looking at the diagram, a brief refresher on ER notation: Entities (rectangular boxes in the diagram) represent “things” or “events”. Interesting relationships between entities are shown by lines, with “cardinality” indicated by *crow's foot notation*. The foot-end of the symbol means *one or more* while the opposite end means *exactly one*; a circle at the 'top of the foot' indicates optionality.

If it seems like too much detail to include such things as **office** and **work_space** in the logical model, remember that we can decide later where to draw the boundary between what will be implemented explicitly, what peripheral concepts will be captured “in name only” (i.e. through **domain_enumeration**), and what other concepts will be eliminated all together from the physical implementation.

However, part of the exercise is to “find a sensible home” in the new, integrated model for every attribute in the fragmented legacy implementation. For example, **office** and **work_space** were conjured up to permit a normalized representation of attributes found in a number of **tdi** tables. In the long run it’s more productive to include too many things in the logical model than too few.

Rather than drawing a boundary to distinguish ‘core’ from ‘periphery’, entities in the Logical ERD are color coded, as shown in the key at the upper right-hand corner of the Figure. An entity’s color indicates whether it’s proposed for implementation, is already implemented, or is shown only for conceptual clarity.

For purposes of discussion, let’s imagine that an instance of an entity in the logical ER diagram corresponds to a row in a database table of the same name. With that perspective, meanings of the entities are presented following the Figure.

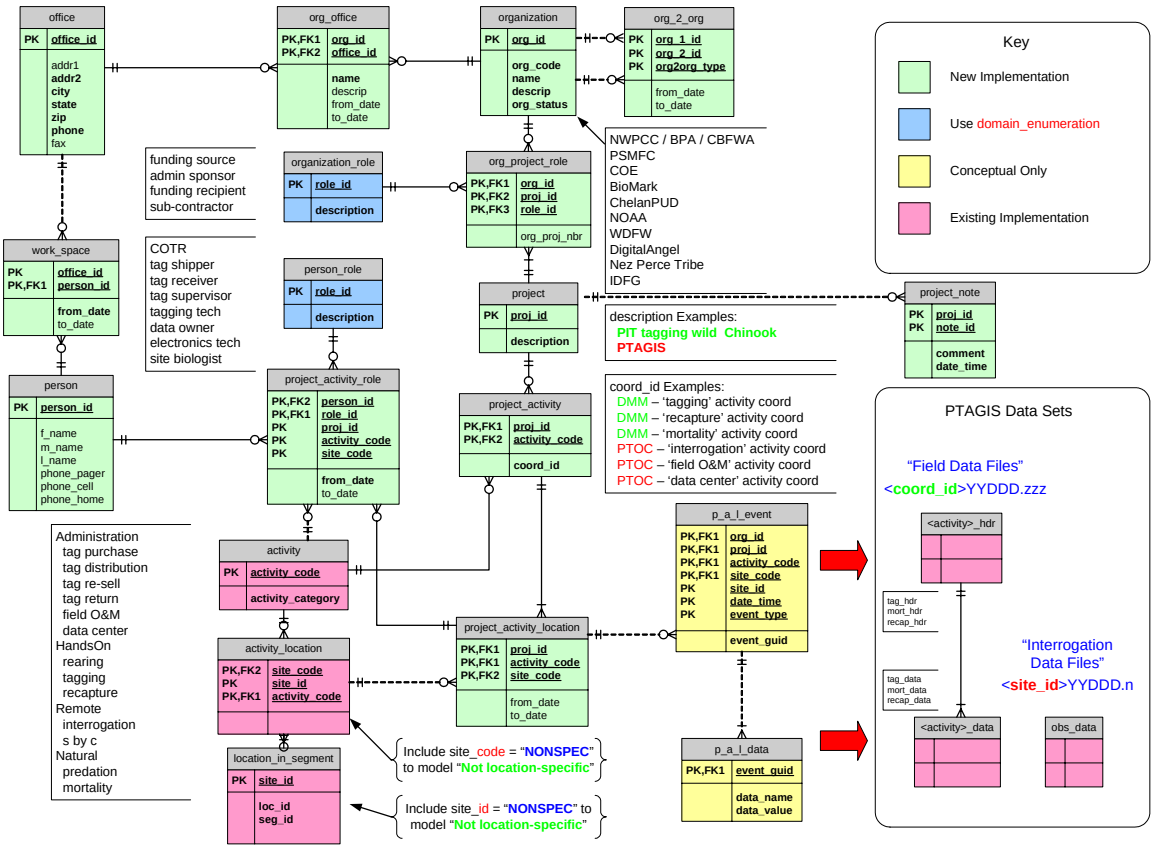

Figure 1 – PPO Logical ERD – Version 1.1

Entity-Type	Description
activity	Names, identifies, and categorizes one of the things a **project** may do, as defined in the ALRS sub-schema. Includes activities performed by the PTAGIS project itself, such as 'tag distribution'.
activity_location	Associates one **activity** with a geographic location, as defined in the ALRS sub-schema. For 'project management', 'tag distribution', or other non-field activities it could be unnecessary to identify a geographic location. Accordingly, special **activity_location** and **location_in_segment** records would be instantiated to represent the "**Not location-specific**" case.
location_in_segment	Associates geographic coordinates with a river segment and describes the principal features of the natural or constructed environment that a naïve observer would see there. Special cases are included for geographic coordinates not located in proximity to a river, and to accommodate circumstances where geographic coordinates are not applicable.

office	Names and identifies a place at which the associated **organization** conducts its business. Attributes might include office-manager e-mail address, switchboard phone number, street address, city, state, and zip-code.
org_office	Associates an organization with one or more offices. Enables re-assignment of an **office** to a new **organization** instance, while preserving the history of previous office-organization relationships, in the event of an organizational name change or a structural re-location of one or more offices between organizations.
org_2_org	Enables the modeling of *temporal* and *hierarchical* relationships between **organization** instances. Refer to PPO_ReOrg_V0.1 (Ver 0.4 Reader Comments on modeling organizational changes) for explanation and examples.
org_project_role	Defines the span of time during which an **organization** exercises a specified **organization_role** with respect to the associated **project**. A typical project has ties to multiple organizations; these are represented by sets of **org_project_role** and **project_activity_role** instances and the associated instances of **person**.
organization	Names and identifies one state, federal, or tribal entity, or agency thereof involved with one or more PTAGIS projects during a stated span of time. Includes PSMFC, as well as commercial vendors (e.g. DigitalAngel) and sub-contractors (e.g. BioMark).
organization_role	A clearly-defined collection of responsibilities capable of being discharged by an **organization**. Unless it is desirable to store additional attributes of **organization_role**, beyond *role_id* (i.e. a name of up to 20 characters) and *description*, it is proposed that **organization_role** be implemented by entries in the existing **ptagis_domain** and **domain_enumeration** database tables.
p_a_l_data	Read "project_activity_location_data": Contains one data element name-value pair pertaining to the associated p_a_l_event. For tagging, release, recapture, and mortality activities, these name-value pairs map into fields of TAGGING file "body records". For interrogation activities, they map into fields of INTERROGATION file "interrogation records".

p_a_l_event	Read "project_activity_location_event": Contains attributes describing an "activity event" along with a GUID identifying one or more p_a_l_data instances. For tagging, release, recapture, and mortality activities, the p_a_l_event attributes map into TAGGING file "header records". For interrogation activities, they map into INTERROGATION file "open" and "close" records.
person	Names and identifies an individual. Attributes might include cell-phone number, pager number, and home phone number. Each logical instance could be implemented by an entry in the LDAP database, the PTAGIS3 database, or both, as determined by business rules yet to be established.
person_role	A clearly-defined collection of responsibilities capable of being discharged by a **person**. Unless it is desirable to store additional attributes of **person_role**, beyond *role_id* (i.e. a name of up to 20 characters) and *description*, it is proposed that **person_role** be implemented by entries in the existing **ptagis_domain** and **domain_enumeration** database tables.
project	Names and identifies a set of project_activity instances, undertaken by one or more organizations at specified locations during a stated span of time, resulting in the creation of one or more PTAGIS data sets. How to identify these data sets requires further thought in the case of 'small-stream interrogation' activities at 'super-sites'.
project_activity	Each instance corresponds to one **activity** conducted by the associated **project.** The **coord_id** attribute contains the three-character code identifying the individual responsible for data files generated by this activity. For example, in the case of tagging, this would be the 'tagging coordinator'. Note that a TAGGING file may contain tagging, recapture, or mortality data - or any combination of the three. So, implicitly, any project engaged in tagging has the capability to engage also in recovery of morts or live fish. Under this model, a **project_activity** would be instantiated for every **activity** the **project** undertakes; specifically, tagging, recapture, recovery, mortality - and interrogation, in the case of a small-stream project. **coord_id** values could differ or could be the same for the various activities.

project_activity_location	Defines the span of time during which the associated **project** makes use of a PTAGIS **activity_location**. Note that in this version of the ERD, for main-stem PIT-tag interrogation activities the *proj_id* does not necessarily match the *site_code*. In other words, a given proj_id could encompass many site_codes - or none, as modeled by the NONSPEC - "Not location-specific" - site_code value.
project_activity_role	Defines the span of time during which a **person** exercises a specified **role** with respect to the associated **project_activity_location**. Note that 'data owner' is a **role** that could be assigned to a **person** responsible for maintaining the collection of tagging files generated by a given **project**. When a project generates data files under multiple activities, a single person could be designated as 'data owner' for all of the files; or this responsibility could be divided among several people. The model would support multiple people being assigned the same **role** for a given **project_activity** during the same span of time. Business rules must be established to determine the circumstances in which this scenario would be allowed.
project_note	Records a textual comment regarding the project.
work_space	Identifies the physical space within an **office** occupied by the associated **person** during a stated span of time. Attributes might include office number, mail-stop, phone extension, direct-dial phone number, personal e-mail address. Note that, at any given time, a **person** could have multiple **work_spaces**, in **office**s of more than one **organization**. Further, it would be possible for a person affiliated with 'Organization A' through **work_space** and **office** to exercise a **project_activity_role** on a **project** 'belonging to' 'Organization B'. For example, a BioMark employee could be identified as ‘Site Coordinator’ (or “interrogation activity coordinator”) at an interrogation facility owned and operated by Chelan PUD.

---------- END ----------