California Franchise Tax Board (FTB)
Enterprise Architecture Council

Enterprise Architecture Definition
Data Management and Delivery

Version No. 1.0

April 11, 2008

Author:
Enterprise Architecture Council

FTB Proprietary and Confidential

Document Information

Document Source

This document is controlled through Document and Deliverable Management. To verify that this document is the latest version, contact Enterprise Architecture.

Revision History

Version No.	Date	Summary of Changes	Revision Marks

Table of Contents

1.0 Executive Summary and Charter 6

1.1 Overview 6

1.2 Scope 6

1.3 High-level Requirements 6

1.4 Conceptual Model 7

1.4.1 FTB Enterprise Business Information Model 7

1.4.2 Enterprise Technical Architecture: Data Context 9

2.0 Current Capabilities and Components 11

2.1 Current Governance and Organization 13

3.0 Target Capabilities and Components 14

3.1 Data Architecture Principles 14

3.2 Common Data Architecture 14

3.3 Physical Data Architecture 16

3.3.1 Target Data Architecture 16

3.3.2 Third Party Data 17

3.3.3 Data Formats 18

3.3.4 Transformation and Validation Processes 18

3.4 Data Standards 19

3.5 Metadata 19

3.6 Data Quality 20

3.7 Technical Architecture 20

4.0 Target Enterprise Governance 21

4.1 Target Data Governance 21

4.2 Responsibility of the Data Center of Excellence to the Business 22

5.0 Gap Analysis 24

5.1 Gap Analysis Defined 24

5.1.1 Strategies 26

5.2 Roadmap 28

5.2.1 Phase Recommendation 28

6.0 Appendix 31

6.1 General Definitions 31

6.2 FTB Data Principles 33

6.3 FTB Enterprise Data Architecture *(Subject Areas and Data Entities)* 37

6.3.1 Subject Areas 37
6.3.1 Data Entities 46

6.4 Industry Best Practices 55

6.4.1 Data Management Body of Knowledge (DMBOK). 55

6.5 Industry Trends 60

6.5.1 Database Federation 60
6.5.2 Operational BI 60
6.5.3 XML and Data Services 60

List of Figures

Figure 1.3-1: Data Management and Delivery Architecture Definition – Integrated Requirements 6
Figure 1.4-2: Data Management and Delivery Architecture Definition – Target Technical Architecture Model 9
Figure 1.4-1: FTB Systems, Database Management Systems and Operating Systems 11
Figure 2.1-1: Current FTB Data Governance 13
Figure 3.2-1: FTB's Primary Data Subject Areas 15
Figure 3.3-1: FTB's Target Physical Data Architecture 16
Figure 3.3-2: Target 3rd Party Data Acquisition and Data Quality Architecture Model 18
Figure 4.1-1: Target FTB Data Governance 21
Figure 5.1-1: Gap Analysis Table 24
Figure 5.1-2: *3rd* Party Data Target Process Model Descriptions 26
Figure 5.2-1: Data Management and Delivery Architecture Definition Phases 28
Figure 6.3-2: Party Subject Area 38
Figure 6.3-3: Customer Subject Area 39
Figure 6.3-4: Asset and Income Subject Area 40
Figure 6.3-5: Tax Declaration Subject Area 42
Figure 6.3-6: Customer Account Subject Area 43
Figure 6.3-7: Human Resources Subject Area 44
Figure 6.3-8: Business Infrastructure Subject Area 45
Figure 6.4-1: Data Management Associations (DAMA) Data Management Framework 55

1.0 Executive Summary and Charter

1.1 Overview

Data Management and Delivery (DMD) comprises all the disciplines related to managing data as FTB's most valuable resource second to its employees. It is the development and execution of architectures, policies, practices, and procedures that properly manage the full data life cycle needs of an enterprise. DMD includes the effective utilization, quality, availability, integration, and standardization of data and information across the enterprise. It defines the management, acquisition, cleansing, storage, and utilization of all data.

1.2 Scope

The scope of the DMD Architecture Definition includes the storage mechanisms, data architecture and structures, data quality, data standards, and administration. The DMD Architecture Definition deals with data residing within FTB systems. Out of scope for this document are Business Intelligence Architecture Definition which includes analytic processing technologies that require Data Management and Delivery to provide their services of quality data; and the Data Exchange Architecture Definition is associated with data from external interfaces, system-to-system integration with external systems, and data that flows through these external facing interfaces.

1.3 High-level Requirements

The following table outlines the high level requirements of the DMD Architecture Definition.

Figure 1.3-1: Data Management and Delivery Architecture Definition – Integrated Requirements

Requirement	Description
Standards Based	The services must conform to widely adopted data standards and the FTB Business Information Models and Structures.
Data Quality	Data quality practices and controls are required throughout FTB's network of applications and data sources to ensure the data is complete, accurate, consistent, and timely. Data quality standards must ensure current state and target state data meets FTB's business needs and process standards.
Data Sharing	Data sharing is an integral part of this Architecture Definition. Data sharing is enabled primarily by the use of query and transaction oriented data services. Data services are predominantly web services that expose data from information to the enterprise, using open standards so that any potential client can use it.
Data Profiling	Data profiling processes check and improve the quality of data. Data profiling procedures, standards, and tools are specified so that quality assurance may be carried out periodically and on demand. The capabilities of the tools and the users of the tools should be maximized. Purchase of tools or better use of existing tools is required.
Data Cleansing	Data cleansing is necessary and performed at an enterprise level. These services are highly scalable and capable of handling the large volumes of data processed by FTB.
Data Transformation	Data transformation is required when the data types of an entity or attribute do not match between the source and destination.
System Performance	The hardware and infrastructure must meet the business performance standards for responsiveness and meet the enterprise needs for efficient handling and processing of data.

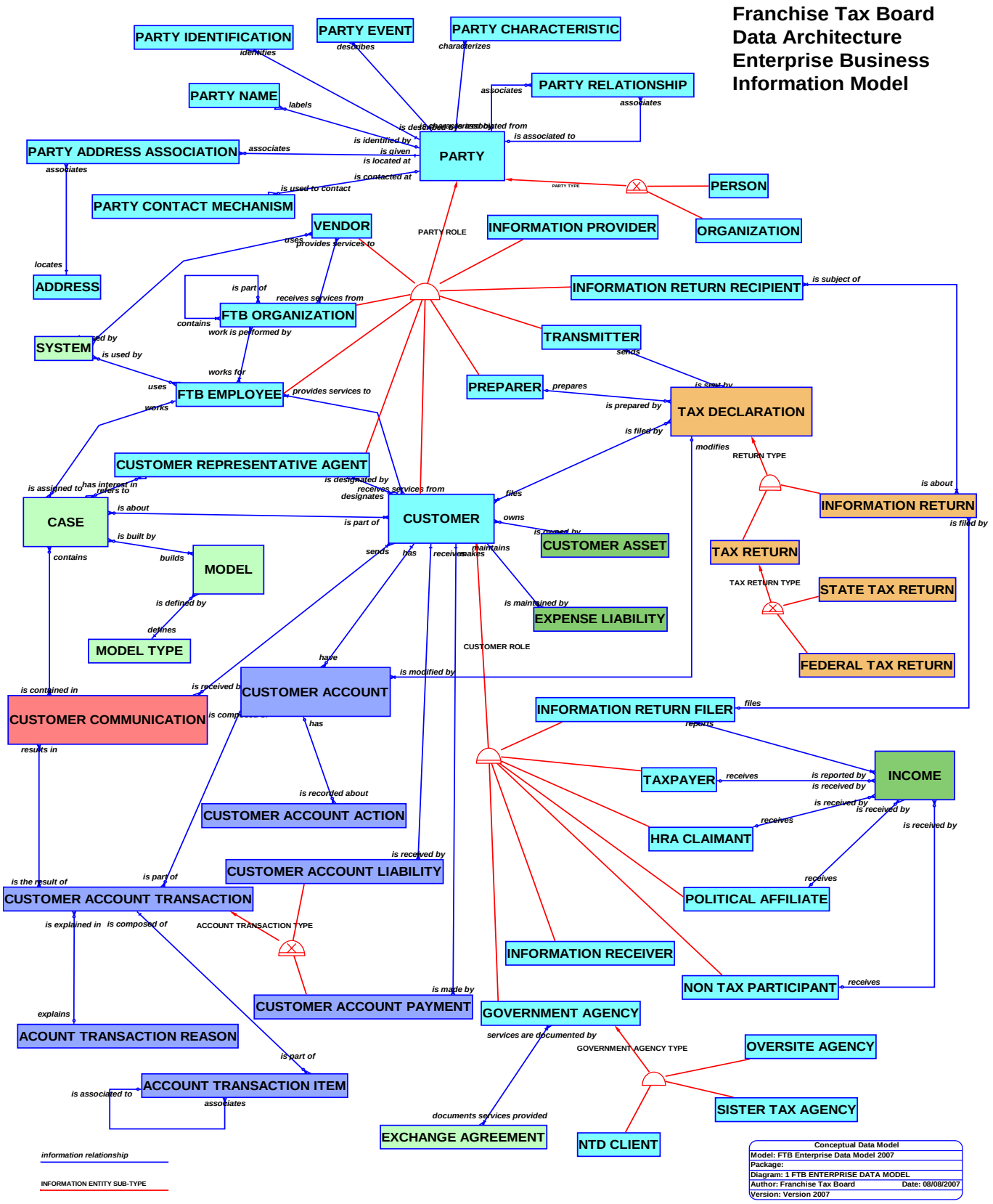
Requirement	Description
Comprehensive federation or integration of data provided through services	The data for taxpayer folder should represent a comprehensive view of a taxpayer. This will require integration of many systems with delivery through data services. All taxpayer related information must be, at a minimum, associated and available by taxpayer through federation or integration.
Security and Audit	Security of data and audit of use are required. The source of the data and the particular requirements for disclosure of the data are managed according to security and disclosure policies.
Abstraction	The management, consolidation, or movement of data in internal systems must be abstracted by the use of services that maintain a consistency for calling applications.
Administration	The technical architecture and system performance must provide efficient management, consolidation, or movement of data in internal systems.

1.4 Conceptual Model

1.4.1 FTB Enterprise Business Information Model

The figure below is the FTB Enterprise Business Information Model that contains the diagrams depicting major information categories that support core business areas. It is an integrated model of all information showing each subject area. This diagram shows the major areas of interest to FTB and their relationships to each other. These relationships represent important business rules that describe how we do business. The model aids understanding of data relationships across the department. See Appendix 3.3 for further information.

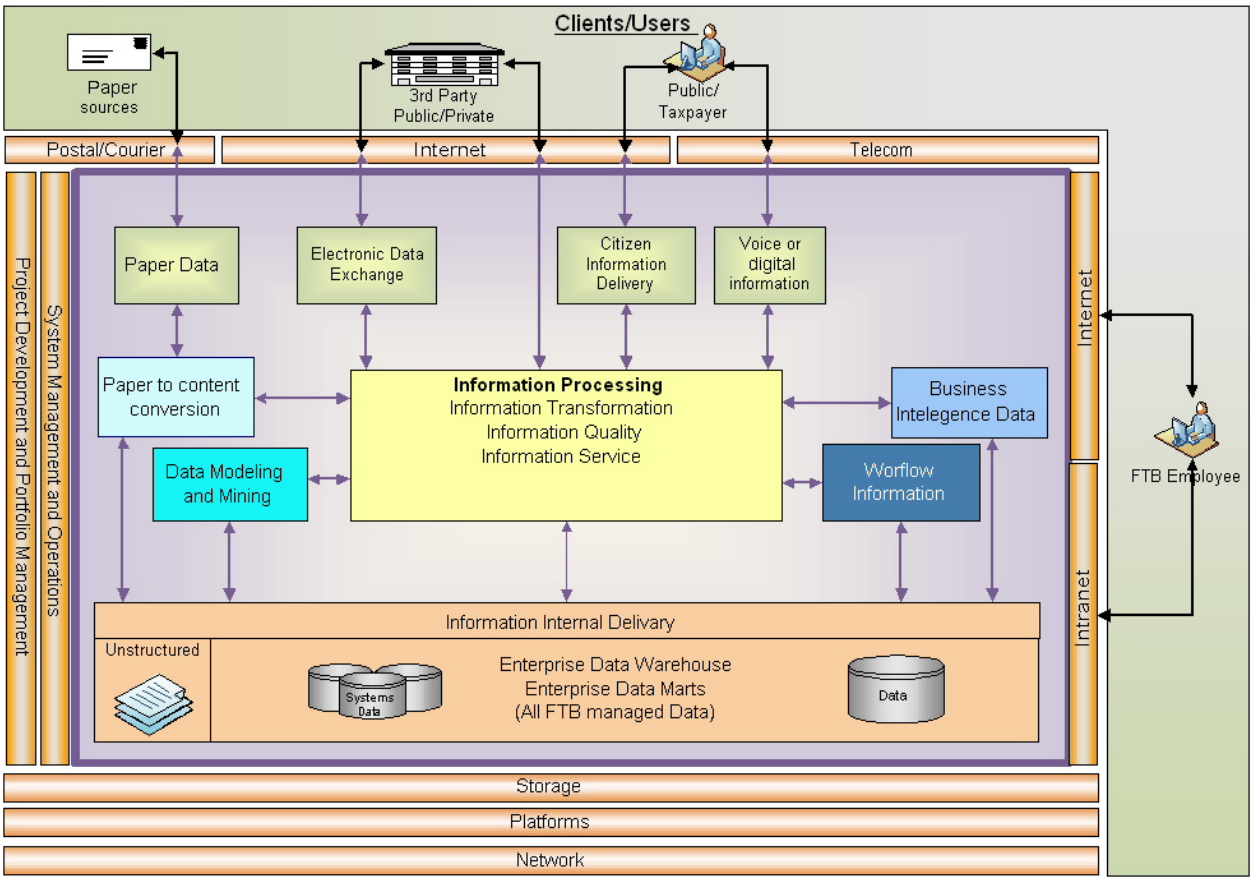
Figure 1.4-1: FTB Enterprise Business Information Model



1.4.2 Enterprise Technical Architecture: Data Context

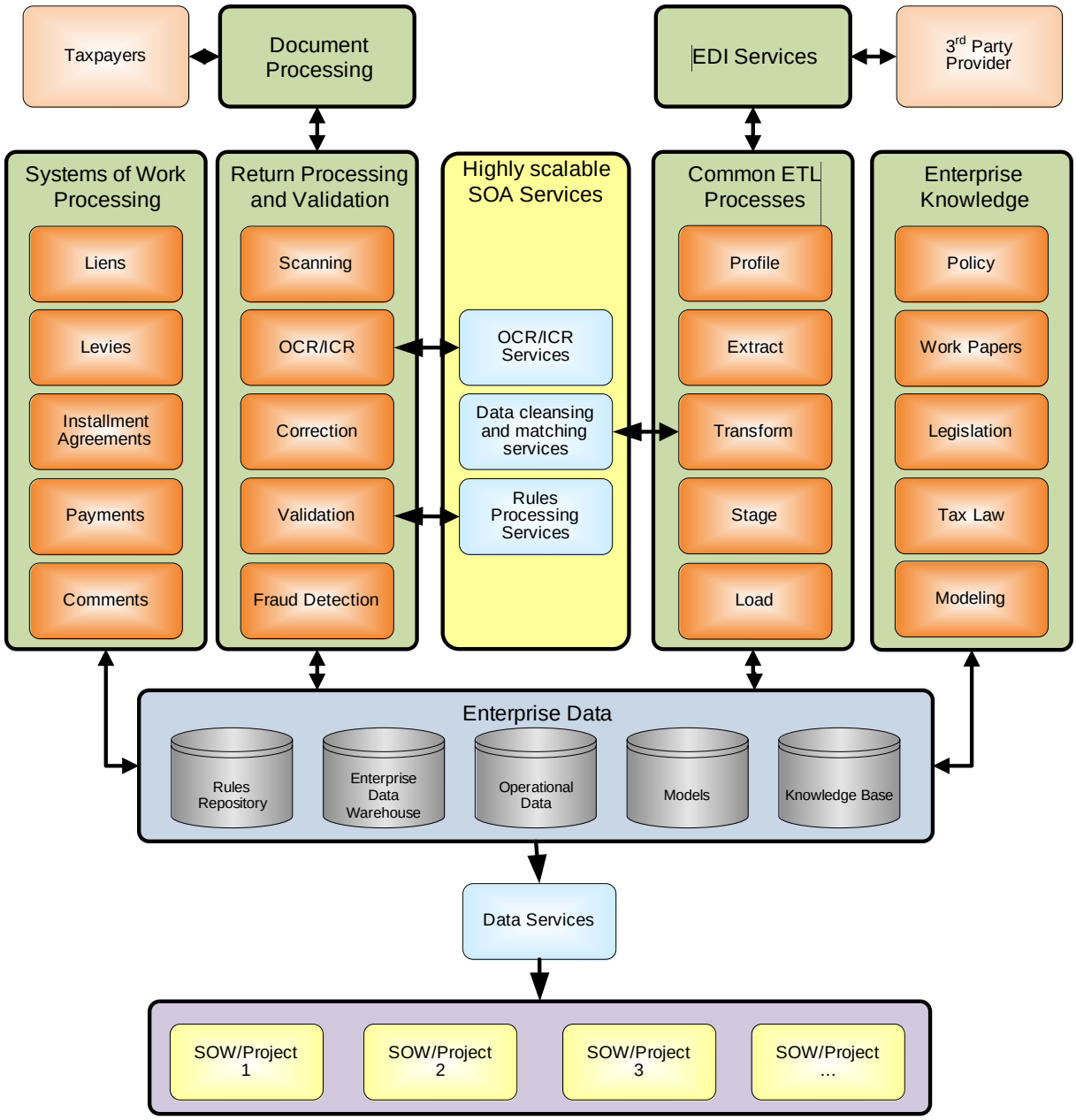
The figure below shows functionality relevant to data at an enterprise level. The data for Taxpayer Folder, 3rd Party Data, and Audit are shown with relationships and interactions between components illustrated using green arrows.

Figure 1.4-2: Data Management and Delivery Architecture Definition – Target Technical Architecture Model



At an enterprise level, there is a need to acquire more data, manage data more efficiently, share data more widely, and deliver data in an efficient manner in order to provide information to enable business opportunities. Some vital components are to improve data exchange, acquire new data sources, extract data from paper forms, build an enterprise data warehouse, and expose data through data services. These tasks require the right technical architecture, tools, services, and business drivers to accomplish these efforts. These are illustrated in the following figure.

Figure 1.4-3: Data Management and Delivery Architecture Definition – Vital Components



Reusable data services are the means of consolidating and sharing data from across the enterprise to the SOWs. These services will present data in a manner that best suits the needs of the enterprise and in a consistent form while still allowing for management, consolidation, and movement of the data in internal systems. Service Oriented Architecture (SOA) implies reuse and leveraging of services by more than one user for more than one purpose. The disparity of the types of databases that exist present some challenges. There are two types of data services possible, the “query oriented” data services and the “transaction oriented” data services.

2.0 Current Capabilities and Components

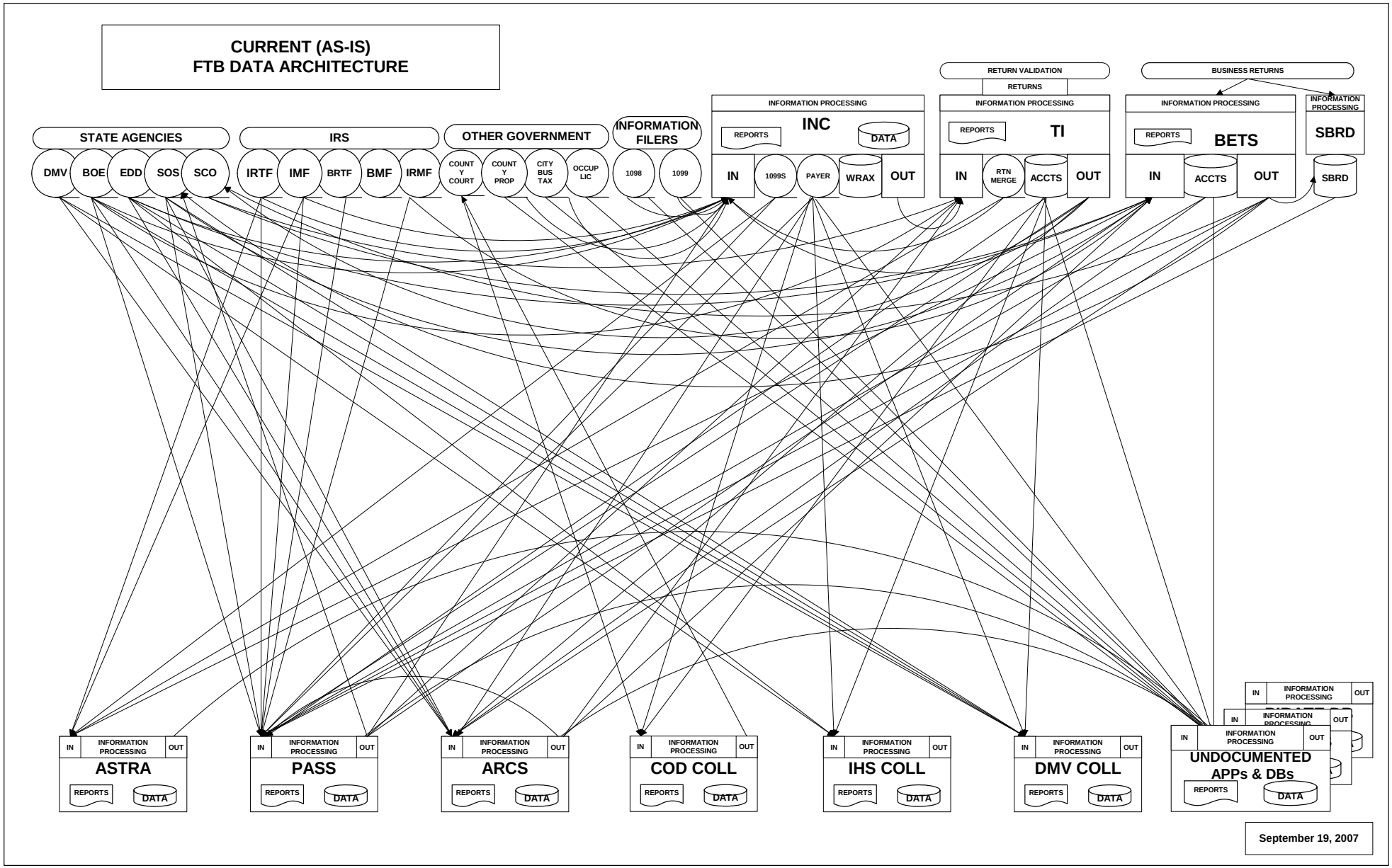
Systems at FTB were developed to meet a specific business goal, and were developed as stovepipe based systems. Multiple operational data stores exist across many database management systems and on multiple operating system platforms. Complex series of batch programs pass data between these systems nightly. There is a small amount of real time data sharing accruing by utilizing various homegrown and purchased middleware products. Common data is replicated across systems. The following table is a partial list of FTB systems along with the Database Management System and Operating System that each system primarily uses.

Figure 1.4-1: FTB Systems, Database Management Systems and Operating Systems

	Adabas zOS	DB2 UNIX	DB2 zOS	FoxPro Windows	MS SQL Windows	Oracle Windows	Sybase UNIX
PIT ARCS	x						x
BE ARCS			x				x
COD	x						
VRC	x						
IHS				x			
TI	x						
BETS			x				
INC		x					
IPACS						x	
IDAX						x	
POA	x						
EWBS							x
EWLS							x
RV	x						
eGateway					x		
ESS					x		
SEC					x		
Astra	x						
Fedstar					x		
PASS							x
Stars	x						
NRWS					x		
Wagwhvr	x		x				x
EAMS					x		
TimePortal					x		
ARMR					x		
ECAIR		x					
PASS MI					x		
PIT Return					x		

The following figure illustrates FTB's current Data Architecture. Major operational data received from outside sources is shown as flows to the systems that process and use it. This figure illustrates the magnitude and complexity of our current data exchanges; it is not intended to be all encompassing.

Figure 1.4-2: Current FTB Data Architecture



2.1 Current Governance and Organization

The current data management and delivery functions are governed by FTB's Data Architecture, Database Support Section, and the system/application project leads. There is no formal governance for data management and delivery. The following figure illustrates our current data management functions.

Figure 2.1-1: Current FTB Data Governance

Entity	Function
Data Architect	• Maintains Enterprise Business Information Model of FTB's core business data architecture. • Performs Data Needs Analysis services with the primary and downstream business communities (logical data modeling) which is mandatory for large projects, and voluntary for small projects. • Reviews data models for compliance of data architecture. • Maintains Data Naming Standards which is used as guidelines and voluntary.
Database Administrator	• Performs database performance and maintenance services. • Performs physical data modeling services. • Installs and supports Database Software and Data inquiry products. • Performs ongoing product support and training, to include a regular forum for problem resolution and information exchange. • Ensures a transparent and productive data environment to support the FTB mission.
Project Architect/Lead	• Develop physical data structures for a project without guidance from Data Architecture and/or the Database Support Section. • Work with immediate business area for a silo data solution.
Business Knowledge Worker (Ungoverned Data Development)	• Many business units create databases or spreadsheets that contain mission critical and secure data. • Data is used for a quick turnaround that cannot be developed by traditional IT services for reports and processing that is internal and external to the department. • Ungoverned databases and sophisticated macro imbedded spreadsheets are sometimes turned into applications with non-secure datastores for use in one or more business units.

3.0 Target Capabilities and Components

3.1 Data Architecture Principles

In support of FTB's Data Architecture, principles for the ongoing management and use of data as a strategic resource have been developed in the five major areas listed below:

Unbold – difficult to read

1) **Information is a strategic business resource "owned" by the enterprise.**

2) **Data must be shared, integrated and accessible across the department.**

3) **Data quality must be maximized to provide best value to the business.**

4) **System solutions must be architected.**

5) **Applications must be partitioned from information processing and management reporting.**

Each principle is fully described in Appendix 6.2.

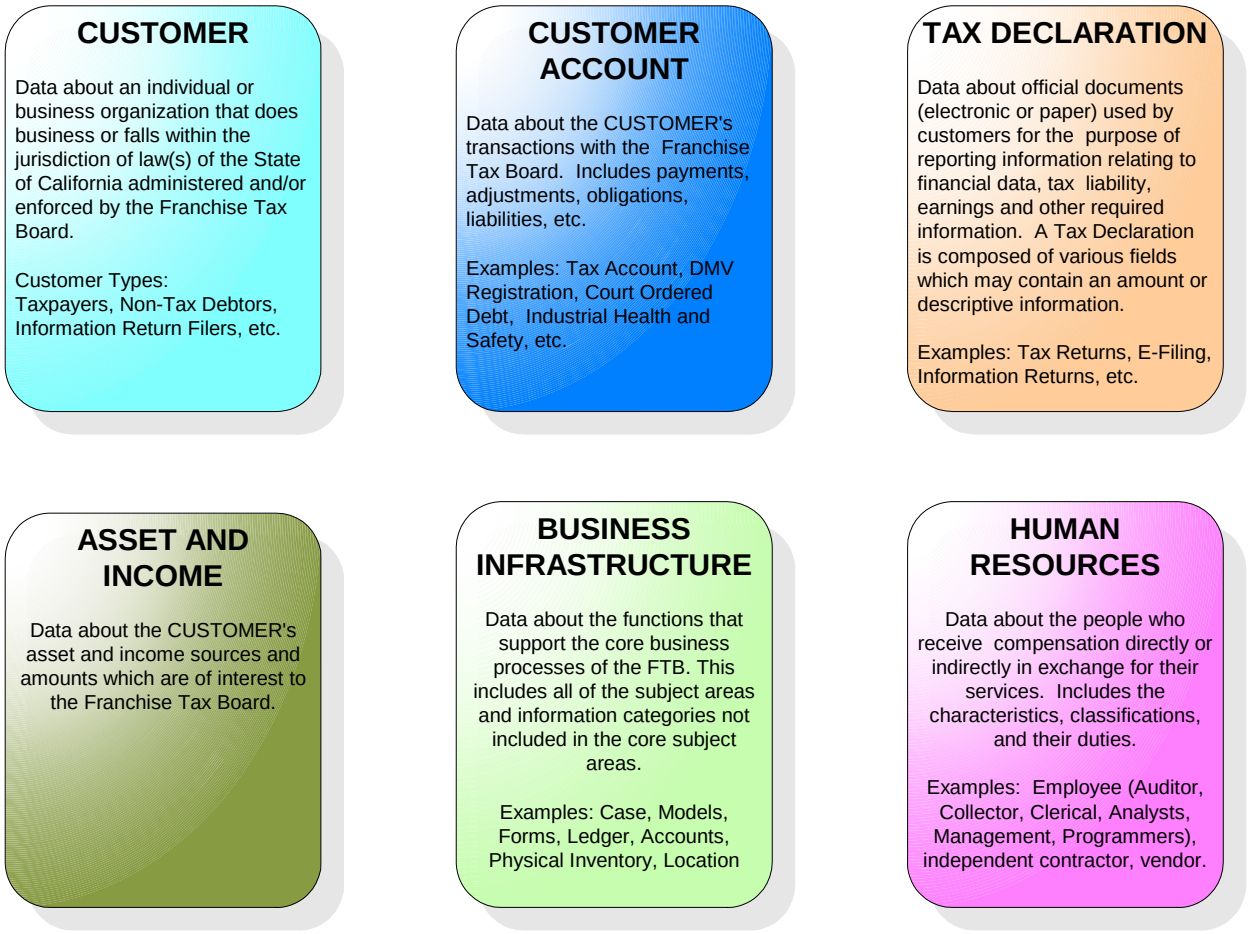
3.2 Common Data Architecture

FTB's corporate data resource will be developed within a single, organization wide common data architecture. Common data architecture is the science and method to design and construct a data resource that is business driven, based on real world objects and events as perceived by the organization, and implemented into appropriate operating environments. It is the overall structure of a data resource that provides a consistent foundation across boundaries to provide easily identifiable, readily available, high-quality data to support the business information demand.

The common data architecture is a formal, comprehensive data architecture that provides a common context within which all data disposed is understood and integrated.

The common data architecture will be subject area oriented. It is built from data subjects that represent business objects and business events in the real world that are of interest to us and about data that is captured and maintained. The subject oriented approach is industry preferred opposed to the project oriented approach as found in typical BI data stores. The subject oriented approach prevents siloed data and gives credence to integrated data. The following figure illustrates FTB's primary common data architecture subject areas:

Figure 3.2-1: FTB's Primary Data Subject Areas



The common data architecture contains concepts, principles, and techniques for developing and maintaining formal data names, comprehensive data definitions, proper data structures, precise data integrity rules, and robust data documentation. Documentation about the data resource is often referred to as metadata, which is commonly defined as data about the data.

FTB shall thoroughly document its primary data resources and 3rd Party data[1] resources. Data resource data are any data that document the data resource and help people understand, manage, and use that data resource to support the business information demand.

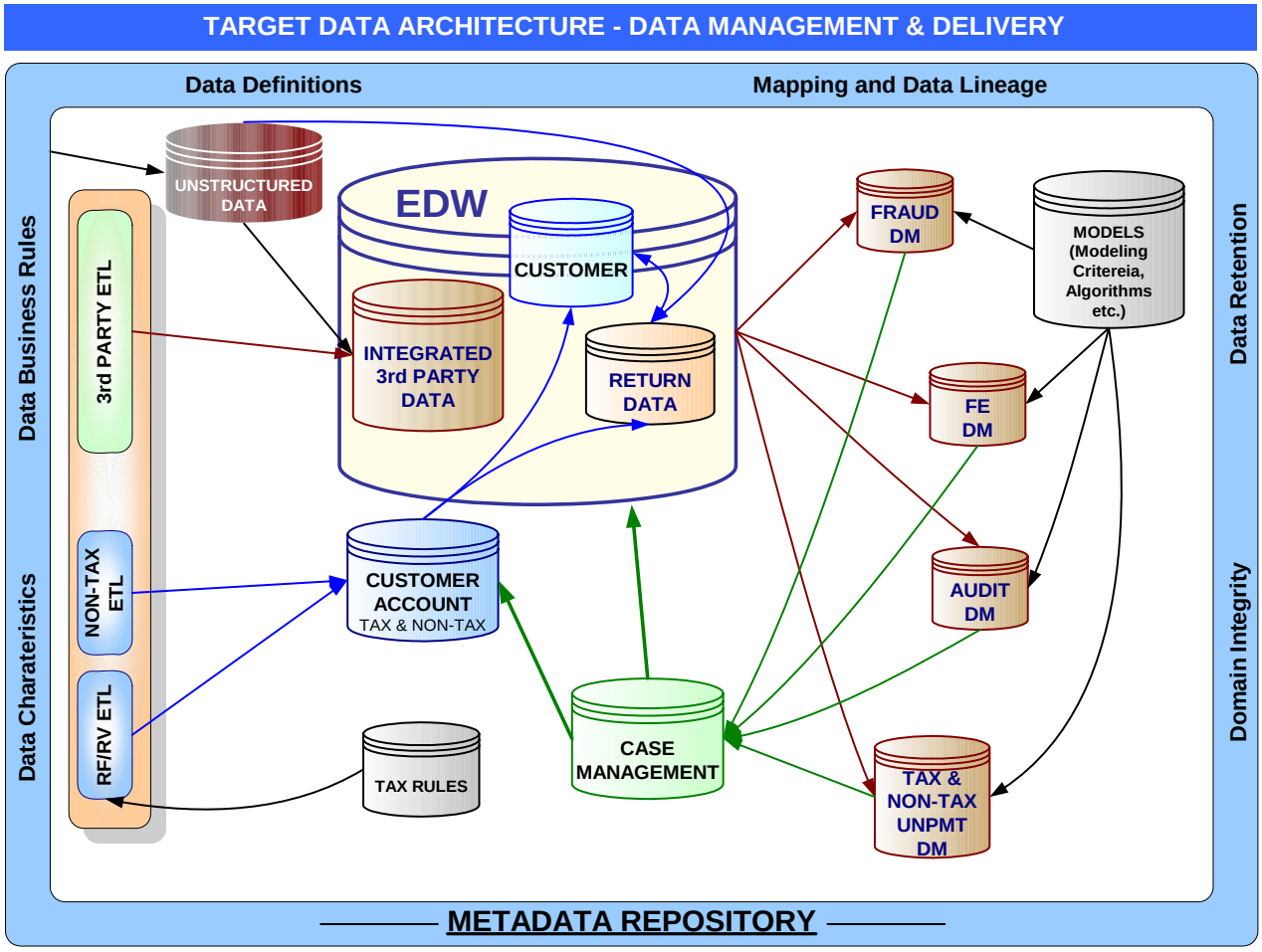
[1] Note that 3rd party data may include restrictions on access or use of the data consistent with state and/or federal law and the agreement with the data source.

3.3 Physical Data Architecture

3.3.1 Target Data Architecture

The figure below illustrates the target physical data architecture. The diagram is a representation of the information flow into the FTB operational and data warehouse data stores.

Figure 3.3-1: FTB's Target Physical Data Architecture



The outer layer shows the metadata repository where all data objects coming into FTB will be cataloged and defined. Unstructured data must be defined and identified with its purpose and author. Data mapping and lineage, impact analysis, data characteristics, data business rules, retention, and domain integrity information will be a part of the metadata repository to improve data quality and the measure of the data quality.

Within the Data Management architecture, the data is extracted, transformed, and loaded (ETL) into data stores. The metadata repository and ETL of data will not be circumvented by manual coding or any other means as this will result in the degradation of data quality and possibly make the data unusable. (See the following 3rd Party Data section)

The customer account subject area data store represents the PIT and BE financial accounting system stores. The Enterprise Data Warehouse contains the 3rd Party data, customer/ customer demographic data, and return data.

The customer data shall be managed in one customer data store. All customer and address data shall be stored in a centralized datastore to eliminate redundancy to avoid creating a new customer database every time a new application is created.

The data marts are created for enterprise modeling and other predictive analytics for identifying non-filers, audit cases, and collections cases. One of the outputs from the data marts feed the case management data store and system. The case management system may be considered as operational and is fed back into the enterprise data warehouse for historical purposes.

The models and tax rules data stores are for recording and storing the criteria/ranges for models and the tax year rules respectively. The models and rules are decoupled from the operational systems so they can be more agile for frequent business needs.

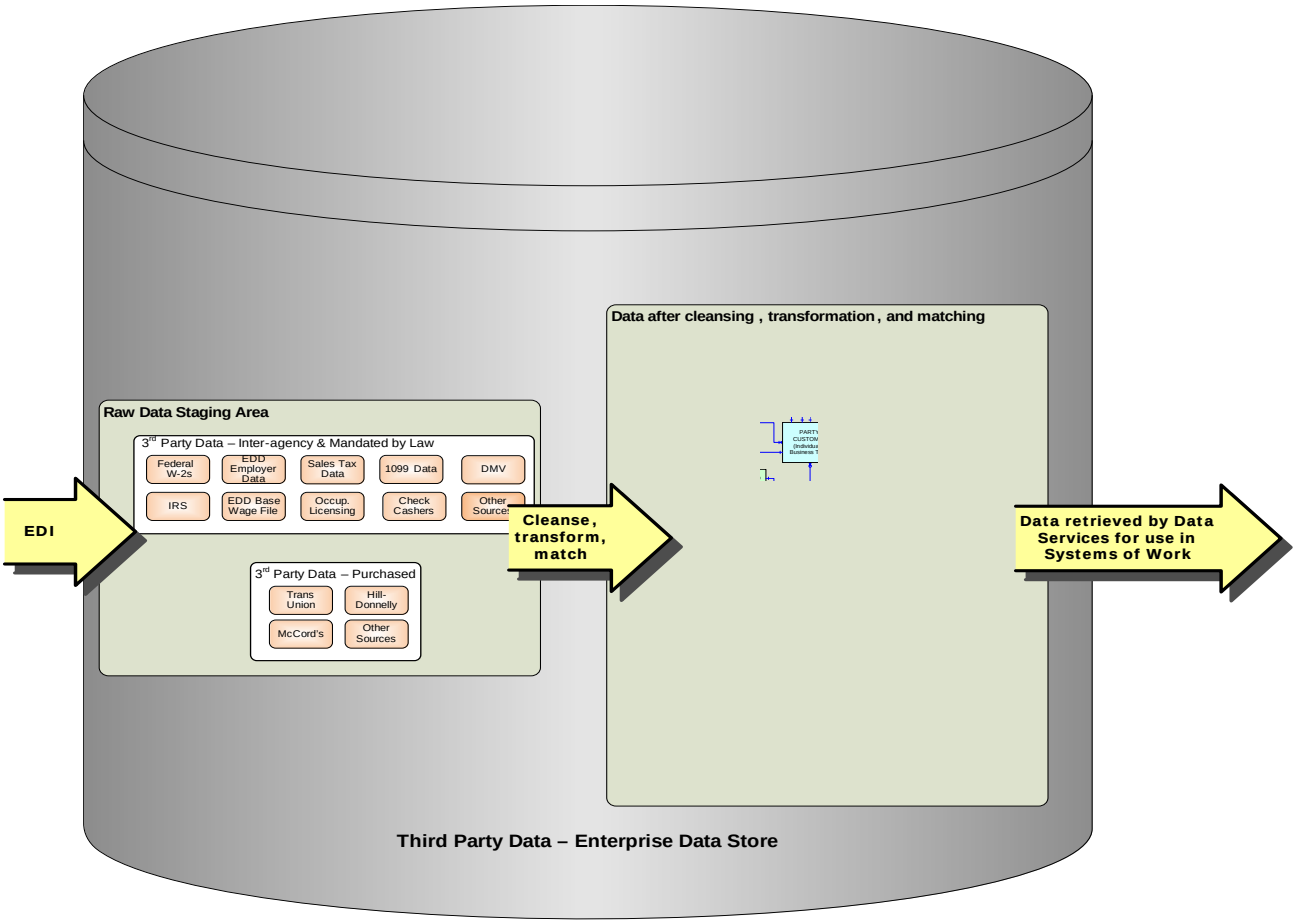
Data Management and Delivery requires the quality of the data delivered be consistent, reliable and timely.

3.3.2 3rd Party Data[2]

The following figure depicts the 3rd Party Data Acquisition Model. The 3rd party data from the data providers is in raw format within a data staging area. The data is ready for cleansing along with the post cleansing and matching data that is stored in a format that is most appropriate for enterprise use. Ideally, the data is obtained by electronic transfer rather than magnetic media or other legacy formats. This data exchange includes documents, spreadsheets, pictures, video, and audio files.

[2] Note that each 3rd party data source may include restrictions on access or use of the data consistent with state and/or federal law and the agreement with the data source.

Figure 3.3-2: Target 3rd Party Data Acquisition and Data Quality Architecture Model



During the data acquisition process, the data will be stored and organized in the form it arrived in, therefore the physical data structures will not be in an entity driven organization. For example, EDD data and DMV data both contain party address data. It is likely the structure of the address data is specific to each provider. As part of the data quality process, data is cleansed, transformed, matched, and copied into a standardized data structure that is ready to be used by the enterprise.

The cleansing, transformation, and matching processes have data exchange components and components that are not related. During the process of cleansing, transforming, and matching, errors may be discovered requiring records to be returned to the sender for correction and resubmission. Error reporting is clearly a data exchange function. Transforming and matching are data management related (i.e., altering the data and associating it for internal purposes).

3.3.3 Data Formats

See the Enterprise Data Exchange Architecture Definition document.

3.3.4 Transformation and Validation Processes

Data exchanges often include processes to transform data from one file format into another for uploading into a database. The processes check for duplicates and errors, create logs, drop

records out for manual intervention, and may return messages of acceptance or rejection. Rejection messages often contain reasons for the file rejection. Before the data is loaded into another system, transformation and validation processes will be run. These processes will be enabled before or after the data transfer depending upon the business requirements.

Considerations of a validation process will:

- **Convert file formats.** Often the export data of one application will not be in the same layout as required by the receiving system. Conversion of the file formats will be required.
- **Validate file layouts.** When a file is transferred an edit program is run to determine if the file is in an acceptable format. If the file passes this validation process, processing continues. If the file fails, it is stopped before moving into the system. With either a pass or fail, the system must return a message with the status of the file and an explanation if the file failed.
- **Prevent duplicates**. Before loading in a new record, it's important to ensure that it doesn't already exist in the destination application. There are many different ways to get duplicate data. For example, the file itself could have been sent twice or records within a file may be repeated. This process will eliminate duplicates and database corruption.
- **Backup and logging**. Regularly scheduled backups and restore procedures will be documented and tested. Transactions will be logged for troubleshooting and security tracking according to user and systems requirements.
- **Provide user interface.** For complex processes, an administrative interface will provide for review of what data will change and be used to resolve errors prior to import.
- **Provide user responses.** Alerts, receipts, and acknowledgments are different methods for responding to a customer action or FTB action. Alerts can be email notices for error process or events. (See EDE Architecture Definition Document)

3.4 Data Standards

FTB shall develop and maintain the following data oriented standards:

- Standard data models
- Standard data naming conventions
- Standard data terms and definitions
- Standard data oriented functions, processes, and practices
- Standard data oriented roles and responsibilities
- Standard data formats
- Standard data oriented deliverables and metrics

3.5 Metadata

Metadata is not just data about data. It includes business rules associated with a piece of data. FTB will use a metadata repository product for the enterprise. See previous figure showing the layer for "Data Business Rules and Metadata."

FTB will build an enterprise metadata model that supports multiple contexts through an extensible model that treats ontology as descriptive.

Enterprise metadata will be set in context. Enterprise wide data (e.g., Enterprise Data Warehouse data) can no longer manage data without contextual metadata. All FTB applications and systems must define the context of the data that is being used and not siloed systems, where the context of data is inherent within the system.

3.6 Data Quality

We will support data quality for both current and future business information demand. Data quality is a measure of how Data Management and Delivery supports the current and the future business information demand of the organization. The Data Management and Delivery cannot support only the current business information demand while sacrificing the future business information demand. Data resource quality is stability across changing business needs and changing technology. Stability across change is the ideal that provides the foundation we need to become an intelligent learning organization.

The issues of data quality and data integration are related. Many data quality issues stem from the inability to identify and match data. We will improve data quality by integrating future systems.

3.7 Technical Architecture

We will implement a common customer identification scheme for the customer and their associated records. The common customer identifier provides the ability to integrate data across the enterprise.

Future mission critical datastores will require 24/7 or near 24/7 availability. Future technical architecture will need to meet this need in addition to addressing security requirements, support data retention policies, and integrate these into a recovery plan.

Security shall be integrated and comply with SAM 5300 and FTB standards.

Operational data integration tools that utilize the Metadata Repository shall be used to support database consolidation, migration, upgrades, and synchronization.

We may use database federation or centralized datastores (see Appendix 3.5.1.1)

We will use Service Oriented Architecture data services for the delivery of data from our physical data stores (i.e., the data layer).

4.0 Target Enterprise Governance

4.1 Target Data Governance

Data Management and Delivery is vital to FTB. Similar to any valuable asset, we recognize that our data assets must be managed. Data management seeks to effectively control and leverage data assets. Data governance shall support the data principles discussed in section 3.1 above.

Data Management and Delivery is a shared and collaborative responsibility between the Data Center of Excellence, business data stewards serving as trustees, and of enterprise data assets, and technical data stewards serving as the expert custodians for these assets.

Figure 4.1-1: Target FTB Data Governance

Entity	Function
Data Architect	• Heads FTB's Data Center of Excellence. • Reviews, approves or disapproves data models for compliance of FTB's data architecture. • Maintains Enterprise Business Information Model of FTB's core business data architecture. • Maintains FTB's Data Standards.
Data Center of Excellence	The Data Center of Excellence is composed of the department's Business and IT data experts responsible for: • Data Governance – planning, supervision and control over data management and use. • Data Architecture, Analysis and Design – data modeling and specification. • Database Management – database design, implementation and support. • Data Security Management – ensuring privacy, confidentiality, and appropriate access consistent with applicable state and/or federal law and the agreement with any third party data source. • Data Quality Management – defining, monitoring and improving data quality. • Reference and Master Data Management – managing trusted versions and replicas. • Data Warehousing and Business Intelligence Management – enabling reporting and analysis. • Document, Record and Content Management – managing data outside of databases. • Meta Data Management – integrating, controlling and providing metadata. • Discussion and Approval of Data Related Activities.
Database Administrator	• Performs database performance and maintenance

Entity	Function
	services. • Performs physical data modeling services. • Installs and supports database software and data inquiry products. • Performs ongoing product support and training, to include a regular forum for problem resolution and information exchange. • Ensures a transparent and productive data environment to support the FTB mission. • Is a part of or communicates with the Data Center of Excellence.
Project Architect/Lead/ Manager	• Develop physical data structures for a project with guidance from Data Architecture and the Data Center of Excellence. • Gets approval from the Data Center of Excellence to continue the data project.
Business Staff	• Works with the Data Center of Excellence to communicate the purpose, data needs, security level of data being stored, and the business reason for creating a the data store inside or outside data governance. • Receives approval or rejection of data activities.

4.2 Responsibility of the Data Center of Excellence to the Business

It is the responsibility of the Data Center of Excellence to the business to ensure that:

- <u>Business user should be able to:</u>
  - Know what data the enterprise manages.
  - Know what the data means.
    - Including calculations and derivations
  - Know where the data is stored.
    - At a minimum the Authoritative Source
  - Know who is allowed to access the data (security).
    - If they are allowed to know this
  - Know how to get the data.
  - Know what can be done with the data (privacy, compliance).
  - Know what decisions have been made about the data.
  - Know who has made decisions about the data.
    - Governance
    - Stewardship
  - Know what quality issues exist with the data.
  - Know who else is interested in the data.
    - Stakeholder community
  - Know who to contact if there are issues with the data.
    - Know what processes exist to resolve issues

- <u>With this knowledge, a business user will be able to:</u>
  - Use the data they rely on to perform their assigned responsibilities.

- Ensure that the data can always be turned into information.
- Participate effectively in stewardship functions that assure the quality, privacy, security, and compliance requirements of the data.

- <u>The enterprise as a whole will benefit as the business user also:</u>
  - Uses the data to increase the efficiency of the enterprise's operations including reengineering of business processes to take advantage of improved data understandability, availability, and quality.
  - Uses the data to meet the enterprise's business goals including adaptation to changing market, regulatory, and other environments. And also agility in responding to new opportunities.
  - Uses the data to mitigate risk in the enterprise including reduction of operational risk inherent in the data itself as data quality improves.

5.0 Gap Analysis

5.1 Gap Analysis Defined

To improve data quality and integration, FTB will utilize a common identifier for all records. We will utilize a federal or statewide identification system, such as REAL ID. If no standard identification system is in place, we will implement one for internal use.

Figure 5.1-1: Gap Analysis Table

DATA AND TECHNICAL DOMAIN	GAP	CHANGE REQUIRED	BENEFITS	RISKS	IMPACT
Common Data Architecture	FTB does have a target data architecture that is generally accepted or is completely ignored.	Establish the Data Center of Excellence. Enforce the target architecture through Data Center of Excellence.	Consistency in the development and deployment of all system solutions.	The risk not to have a compatible architecture will continue to result in poor data quality, complex ETL, higher resource costs, and poor performance.	**High**
Determine use of Enterprise Modeling	No Standardization of Data across data source systems and the third party data sources.	Establish data standards to support enterprise-modeling capabilities. This is done by implementing an enterprise data store for modeling activities.	Data has a common format, type and value set for both data and presentation layers From a semantic standpoint data has a common meaning (income from one source has the same meaning as income from another source so that they can be directly compared in a model) Better data will lead to more effective models that will generate more revenue and reduce costs.	If FTB continues to ignore data standardization in terms of modeling data internally and externally, then there will be continued lack of sharing of data between SOWs.	**High**

DATA AND TECHNICAL DOMAIN	GAP	CHANGE REQUIRED	BENEFITS	RISKS	IMPACT
Enterprise Metadata Repository for OLTP and OLAP	Metadata Repositories are silo and hand code. No metadata standards are adhered to.	FTB needs to govern and create a central enterprise metadata repository that is compatible with industry metadata standards	Consistent and Contextual meaning of data reduces the cost of staff trying to analyze each piece of data for meaning and use each time it is used. If it is defined upfront and made useful, then time and cost is minimized for delivery of the BI product.	Governance and creation of a central or federated metadata repository is imperative in BI. If this is not done costs will remain high in developing BI products to the business.	**High**
Data will be shared, integrated and accessible across functional and organizational boundaries	This is an architecture principle that must be respected throughout the enterprise	Business and IT must be educated on the sharing of data and what it means to each line of business.	All data are available across the enterprise in a consistent format using a catalog directory of this data to find it and procedures for getting the data without extensive IT intervention for the business.	All data being available across the department will require disclosure measures and usage guidelines. This data must not be replicated in multiple areas in FTB. There should be access to the source data store.	**High**
The quality of data will be maximized to provide the best value to the business	This is an architecture principle that must be respected throughout the enterprise	All FTB organizations must be educated on the importance of data quality and “Trusted” data. The Data Center of Excellence and the BI Competency Center are the stewards.	Higher levels of quality data. Quality data reduces the need to clean data on the back end / downstream processes like BI. Data must be cleansed early in the tax processes.	The risk of not providing data quality measures will perpetuate the need to constantly cleanse the data through traditional BI ETL when the data should have been cleansed and formatted within compatible data architecture at the beginning of the operational tax processes. Poor data quality equals higher cost, and lost cases in litigations.	**High**

5.1.1 Strategies

Figure 5.1-2: *3rd* Party[3] Data Target Process Model Descriptions

#	FUNCTION/ PROCESS	DESCRIPTION
	3rd Party Data Acquisition	The Improve Data Timeliness Opportunity #2 is due to the expanded use (i.e., enterprise-wide) of 3rd Party Data. While data is timely today for INC, it is likely that more timely data will be needed for other SOWs.
•	Identify New Sources	There is little room for improvement in the identification of new data sources because FE does an excellent job at identifying and evaluating new sources today.
•	Establish Agreements	Agreements with the 3rd party data providers will need to be established that improve the timeliness of 3rd party data. Each agreement with the 3rd party data providers may include restrictions on the access or use of the data imposed by state and/or federal law.
	Electronic Data Interchange (EDI)	Data exchange with providers could be extended to perform end-to-end processing that includes data acquisition, format, cleansing, matching, and error management. Today SWIFT performs the acquisition process, stores the provider file, and has the capability to perform SOW-specific edits. However, this back-end functionality (i.e., format, cleansing, matching, editing) is not consistency used due to staffing and funding constraints.
•	Acquire Data	3rd party data must be acquired on a timely basis to support future modeling activities. Legacy formats needs to be replaced with electronic transmission where possible to streamline the interface between the provider and FTB.
•	Format	The Electronic Data Interchange (EDI) Opportunity #3 enterprise service should be employed to receive, edit, track use, and manage 3rd party data. Record level rejection is added.
•	Cleanse	The Improve Data Quality Opportunity #2 expands data cleansing of addresses and $ amounts using new automated tools and best practices. Manual cleansing is considered taking into account the cost vs. benefit. The Address enterprise service is utilized to improve the accuracy of record addresses.
•	Error Management	Error records and files are returned to the provider. Provider's quality score is tracked for use in provider management and FE processing.
•	Match	The Improve Match Criteria Opportunity #1 uses multiple addresses and possible other demographics such as date of birth.
	Stage and Load	The stage and load process decouples 3rd Party Data from INC and is focused on improving the overall performance of the technical infrastructure.
•	Stage	New architecture, tools, and methods should be considered for staging.
•	Load To INC	Loading of records into any other database will be an implementation decision. The technical architecture of the 3rd Party Data TBO addresses this approach.
•	Load to ECAIR	ECAIR migrates or is replaced by an enterprise 3rd Party Data store. The 3rd Party Data TBO addresses this approach.
•	Receive Returns Information from TI and BETS	TI or BETS information is not stored as part of the 3rd Party Data. Data services are used to access this information independent of 3rd Party Data functionality.

[3] Note that 3rd party data may include restrictions on access or use of the data consistent with state and/or federal law and the agreement with the data source.

5.2 Roadmap

5.2.1 Phase Recommendation

The following are steps to move from the current silo-based database implementation to an integrated state.

Figure 5.2-1: Data Management and Delivery Architecture Definition Phases

Task Name	Scheduled Work (in hours)
Data Delivery and Management	**6,676**
Data Delivery and Management - PHASE I	**480**
Data Delivery and Management Planning	
Coordinate with other ESOs in Data Governance and Standards Planning	320
Document Data Governance and Standards	80
Approve Data Governance and Standards	40
Implement Data Governance and Standards	40
Data Delivery and Management - PHASE II	**6196**
Establish Enterprise Metadata Repository - Phase II-A	
Establish Metadata Governance Processes	80
Determine Metadata Methodology and Requirements for Gathering Metadata	80
Procure Metadata Repository Tool	160
Gather Metadata and Populate Repository	860
Publish Metadata	80
Milestone - Enterprise Metadata Repository Established	
Establish the Enterprise Data Warehouse (EDW) - Phase II-B	
Enterprise Data Warehouse Planning	
Coordinate EDW Planning with BI and SOA	160
Document EDW Plan	80
Review EDW Plan	40
Enterprise Data Warehouse Analysis	
Determine EDW Requirements	120
Develop EDW Logical Data Model	120
Analyze Requirements and Identify EDW Tables	112
Milestone - Agreement on EDW Data Contents	
Enterprise Data Warehouse Design	
Determine Design Approach	96
Document Design Approach	80
Review Design Approach	40
Enterprise Data Warehouse Data Model Design	
Analyze Requirements and Identify EDW Tables	80
Design EDW Views Template	80
Design EDW Physical Data Model	40
Enterprise Data Warehouse Population Design	
Validate Data Replication	24
Design EDW Population Approach	80
Document EDW Population Approach	40
Review EDW Population Approach	16
Design EDW Population Template	80
Review EDW Population Template	16

Enterprise Data Warehouse Data Services Design	
Design EDW Data Service Approach	80
Document EDW Data Service Approach	40
Review EDW Data Service Approach	16
Design EDW Data Service Template	80
Review EDW Data Service Template	16
Milestone - EDW Design Complete	
Integrate Metadata Repository with the Enterprise Data Warehouse	
Integrate and link MDR with the EDW	80
Enterprise Data Warehouse Build and Test	
Build Automation Script for generating View code	80
Confirm EDW Environment	40
Enterprise Data Warehouse Data Structures Build and Test	
Build EDW Database	80
Build EDW Views	80
Enterprise Data Warehouse Data Services Build and Test	
Develop, Test and Review Data Services	360
Enterprise Data Warehouse Population Build and Test	
Develop, Test and Review Replication Process	360
Develop, Test and Review Population Processes	360
Milestone - EDW Build and Test Complete	
Populate Data Marts for Enterprise Modeling - Phase II-C	
Underpayment Model Population	
Support BI in the analysis design and implementation of Underpayment Modeling Data Mart (See BI ESO)	320
Populate Underpayment Modeling	80
Filing Enforcement Model Population	
Support BI in the analysis design and implementation of FE Modeling Data Mart (See BI ESO)	320
Populate FE Modeling	80
Audit Model Population	
Support BI in the analysis design and implementation of Audit Modeling Data Mart (See BI ESO)	320
Populate Audit Modeling	80
Fraud Model Population	
Support BI in the analysis design and implementation of Fraud Modeling Data Mart and OpBI (See BI ESO)	640
Populate Fraud Modeling	120
Data Delivery and Management - PHASE III	**1496**
Utilize Metadata Repository	
Link Metadata Repository to all data items	200
Establish Data Access Through Data/Web Services	
Develop Data Access Requirements for Data/Web Services	120
Review Data/Web Services	40
Verify Data/Web Services Requirements have been met	40
Test Data/Web Services for Performance	120
Approve Data/Web Services	32
Migrate "As-Is" Databases to Target Database Structures	
Data Migration Planning	
Develop Data Migration Plan	80
Document Data Migration Plan	40
Review Data Migration Plan	16
Data Migration Requirements	
Develop Data Migration Requirements	80
Document Data Migration Requirements	40
Review Data Migration Requirements	16

Data Migration Design	
Develop Data Migration lineage and flows from source to target database environment	80
Document the data migration lineage and flows from source to target database environment	40
Review and Affirm the Data Migration Design	32
Data Migration Build, Test and Review	
Build automation scripts for data migration	200
Develop, Test and Review data migration from source to target databases	320
Milestone - Migrate "As-Is" Databases to Target Database Structures	

6.0 Appendix

6.1 General Definitions

Categories of Data - Logical classification of data elements (attributes) that pertain directly to a function or a major topic of interest to the enterprise. (Also see Subject Areas)

Context Diagram - A high-level data flow diagram depicting only the inputs and outputs of a system.

Data Dictionary - A vendor supplied or custom developed database that typically track user populated data element definitions and their occurrences across an organization's applications.

Data Entity - An item that represents a class of People, Objects, Locations, or Events (POLE) having characteristics of interest to the enterprise about which data could be kept. Examples include: 1) People: TAXPAYER, PREPARER; 2) Object: RETURN, CUSTOMER ACCOUNT; 3) Location: RECEIVING, STORAGE LOCATION, ORGANIZATION; 4) Event: CUSTOMER ACCOUNT ACTION.

Data Entity Model / Data Entity Relationship Model - A structured representation of data and its inherent relationships, which depict the functional use of data without regard to its physical implementation. (See Data Model)

Data Entity Type - A collection of data entities to which a specific definition applies.

Data Model - A structured representation of data and its inherent relationships, which depict the functional use of data.

- **Conceptual/Semantic Model** - Contains Data Entities and their Relationships that describe business data concepts despite its physical implementation
- **"First-Cut" Logical Model** - Scoped to the project requirements and includes all the necessary Data Entities, Relationships and Attributes yet without the consideration of physical implementation.
- **"Normalized" Logical Model** - The further decomposition of the data model which includes elimination of repeating data groupings, redundant key values or rows, and attributes dependent upon other non-key attributes
- **Physical Data Model -** A representation of an application specific implementation of logical data model requirements.

Data Quality - Data Quality is accurate and consistent data resulting in useable information and knowledge

Data Sharing - The situation where one type of data is used to support more than one business activity.

Data Steward - Someone who manages data for someone else (i.e., a trustee/custodian). Further, data is not owned by any one person or unit, but owned by the enterprise. Therefore, this person is entrusted by the enterprise with information others require.

Enterprise Model - An integrated view of multiple applications, or application perspectives. Generally speaking, the Enterprise Model represents the "bigger picture" than a single application, function, or database.

Enterprise Pool - See **Subject Area.**

Instance Data - Repository data is considered only as data depending on the perspective of the user. One user's perspective may be considered as metadata by another user.

Logical to Physical Mapping - The identification of logical data model requirements that will be used in the physical data models. Additionally, logical to physical mapping may utilize existing physical data structures, which have been identified in a logical data model.

Mechanism - An automated method or system by which specific processes are executed.

Metadata - A descriptive detail of data instances, also known as "the data about data." Examples of metadata include field names, lengths, program names, definitions, etc., which are usually the detailed description of data resources and business systems.

Metamodel - An encyclopedia/repository model representing a distinct view of the repository information. Examples of metamodels may include the detailed data entities/relationships of a subject area, data flow models detailing a context diagram, etc.

Multityping - A data model where a real world data entity may be described by more than one data entity. For example, CUSTOMER might be classified as a TAXPAYER, INFORMATION RETURN FILER and a NON-TAX DEBTOR. Each classification adds more attributes, relationships, and rules to the knowledge we can express about that CUSTOMER.

Referential Integrity - Create, update, and delete constraints between to data entities/tables.

Repository - An integrated access point for globally resident information represented as metadata (detailed description of data resources and business systems). Repository contents are definable, loadable, and retrievable regardless of the origination tool, platform, programming language, or DBMS. The repository may be thought of as the card catalog in a public library.

Subject Area - A classification or grouping of data entities that pertain directly to a function or a major topic of interest to the enterprise. Examples include: CUSTOMER, RETURN, HUMAN RESOURCES, and CUSTOMER ACCOUNT.

Subtype - A subordinate data entity type in a subtype set. Common attribute and relationship types are "packaged" into the super-type for the subtype set and inherited by the subtypes in the subtype set. In this way, a subtype is a more specialized category of its super-type that inherits all of the super-type’s attributes and relationships, and also may contain additional attributes and relationships that are unique. TAXPAYER is a subtype of CUSTOMER.

Subtype Set - A classification of a super-type's immediate subtypes as exclusive/inclusive and complete/non-complete.

Super-type - A data entity type that is the super-ordinate member of a subtype. A super-type is a generalization of its subtypes. Its attributes and relationship types are inherited by all of its subtypes. CUSTOMER is a super-type.

Trusted Data – Trusted Data comes from a central source (i.e., Enterprise Data Warehouse) where the data has been formatted and verified for quality and use for all downstream processes and datastores.

6.2 FTB Data Principles

1) Information is a strategic business resource "owned" by the enterprise.

All data is an enterprise resource. All organizational units have common ownership of data, although they may have different responsibilities with respect to the data (stewardship). Certain data from third party sources may contain restrictions on the access or permissible use of the data established either by state and/or federal law, or the individual agreement with the data source. The inherent interdependence of organizational units and business processes demands that data be managed at the cross functional or enterprise level.

2) Data must be shared, integrated, and accessible across the department.

Multiple sources of data are combined to form a single enterprise resource. Pools of enterprise data are used by multiple applications and data must be defined and managed consistently across the department. The right data must be available to, and accessible by, the right person at the right time. This means that the definition and location of data must be readily available across the department.

- **Meaningful data standards facilitate understanding, communication, and productivity.** Shared data requires enterprise focused data standards to ensure common understanding and consistency. Data standards exist for two reasons: (1) to improve productivity of the business through common understanding of its information and business rules, and (2) to facilitate communication among business people as well as information systems professionals.
- **To support all customers, data must be named and defined with a singular definition, standard set of values, and uniform business rules.** An example of the business rules are those that include any restrictions on the access or use of the data obtained from third parties that may be imposed by state and/or federal law or the agreement with the data source.
- **The use of already approved data characteristics (such as length, format) is required.**
- **The meaning of data and its business rules must be made accessible to data producers and knowledge workers.**
- **All applications and knowledge workers should access data from a single database unless it is not feasible.**
  It is difficult to assure consistency among distributed copies of data. Replicated data requires a business case and a plan for managing the consistency of the data.
- **Knowledge workers are responsible for the downloaded data they retrieve/receive.** In the same way that employees are accountable for their work, they are accountable for their use of information. Confidentiality or privacy requirements must be adhered to. As recipient of data, the knowledge worker becomes a steward of downloaded data. This does

not become "their" data. If erroneous data is discovered, they should follow guidelines to correct errors at the authorized source, not simply update their copy.

3) Data quality must be maximized to provide best value to the business.

Data quality includes completeness, accuracy, and timeliness. Completeness is the existence of all data needed for business processes; accuracy is the absence of errors; and timeliness means the most current data is available when needed.

- **Capture data as close to its original source as possible.**
    - Data producers should capture data electronically at its source. Data entered by the data producer tends to have the highest level of integrity. Data entered by intermediaries via forms, paper documents or other means, will be more likely to contain errors. Data intermediaries may not have the ability to rectify errors or omissions encountered on the paper form.
    - Data quality is enhanced when data is collected at the original source outside the enterprise.
    - Data recorded by the customer eliminates potential errors encountered when an intermediary or businessperson records customer data.
- **Capture data using the method that provides the best accuracy, completeness, and timeliness.**
  The use of scanners and/or other automated data collection devices may eliminate the potential for human error. Electronic receipt of data or use of scanners can enhance timeliness and completeness of data.
- **Capture data once.**
  Data is the only business resource that is not consumed when it is used (retrieved). Therefore, it need not be reentered for each different use.
- **Those who create/update data are responsible for the quality of the data.**
  One of the basic principles of quality is accountability by persons performing work. Integrity rules cannot be effective without accountability. Accountability requires that data producers be properly informed. Data producers must know their data customers, data usage, and data quality requirements.
- **Define data quality targets; measure and report data quality.**
  To maximize data quality, producers and data intermediaries must know data quality requirements as well as current output quality statistics. Knowledge workers need to know the level of quality in order to use data properly.
- **Use data quality measurements to identify and correct defective processes.**
  If data quality is not acceptable, it indicates that the data collection and evaluation process must be improved. Data producers to measure the quality of their output can use data quality “audits” or assessments. The results of the audits should be used to improve the process.

4) System solutions must be architected.

Systems are business process driven, developed with an enterprise viewpoint, and use data-oriented methodologies. Databases should be designed to facilitate system portability, scalability and adaptability.

- **Redundancy must be planned.**
  Data should be replicated when necessary with one location designated as the source. Replicated data should be synchronized with the source as needed to validate data.

- **Design systems to capture data for downstream business needs.**
  Systems should be architected to capture data for the enterprise and not just for a single business process. Systems that restrict data capture to a single business process are referred to as stovepipe systems.
- **Eliminate redundant data create/update applications, interfaces and data transformations.**
  These redundancies represent the reacquisition of data resources that exist electronically elsewhere. They represent non-value-added costs and put the enterprise at risk for producing conflicting data reports. Capture data once and replicate it as necessary.
- **When unable to use a single data source, distribute data by replicating it, as opposed to interfacing and transforming it to a different format.**
  When data is required in more than one database for timely or cost effective access, it should be replicated or extracted in a common form. The same data fact in multiple databases must have the same definition and set of values to maintain consistency across databases. The only acceptable exceptions are transforms required for approved software packages in which data is not maintainable in the same format. All data transforms require automated integrity controls to assure that data values are consistent.
- **Data quality controls should be independent.**
  Business rules and data integrity controls should be defined and developed separate and apart from the business process logic and data access logic. Because there may be many business events that cause update to data, the integrity logic must be reusable across the applications that can cause change to shared data.
- **Develop applications using single create/update modules around entity types, subtypes and natural business events.**
  Developing applications around the inherent business objects maximize the consistency of data and supports data integrity.
- **Edit/validation systems should be developed separately from operational systems.**
- **Existing routines for data validation, computations and source file access should be used where possible.**
  Existing algorithms for name/id/address validation, etc., should be used where possible, and where not possible, the new ones must follow the same business requirements as previously established.
- **A data impact analysis is an essential part of developing or modifying an automated system, or when changing data definition, location or characteristics.**
  A data impact analysis describes the impact of the new or changed data design on all departmental systems that use or could use the data. This analysis must be performed when access or data usage is impacted by a change in systems or technology. This analysis is used to inform business system managers, current project managers and program managers of the changes so they can accommodate any changes that impact them.
- **Data modeling should be used for data analysis when: 1) developing an application, or 2) making substantial changes to the data base/format of an application.**
  Data modeling (entity relationship diagramming) reveals data interfaces and impacts.
- **Data analysis/data modeling activities should be performed with an enterprise viewpoint to be of value.**
- **The re-use of prior data analysis is strongly suggested.**
- **A logical data model: 1) should be used to architected any new database, 2) should be in synch with the enterprise data model, and 3) should be fully attributed, with each data entity and attribute defined in business terms, including data characteristics.**

- **Architected solutions require the use of enterprise naming conventions, abbreviations and enterprise defined data definitions for any data model or analysis.**
- **The use of a standard CASE tool ensures standardized deliverables. Uniform deliverables facilitate better communication of business information needs throughout the organization.**

5) Applications must be partitioned from information processing and management reporting.

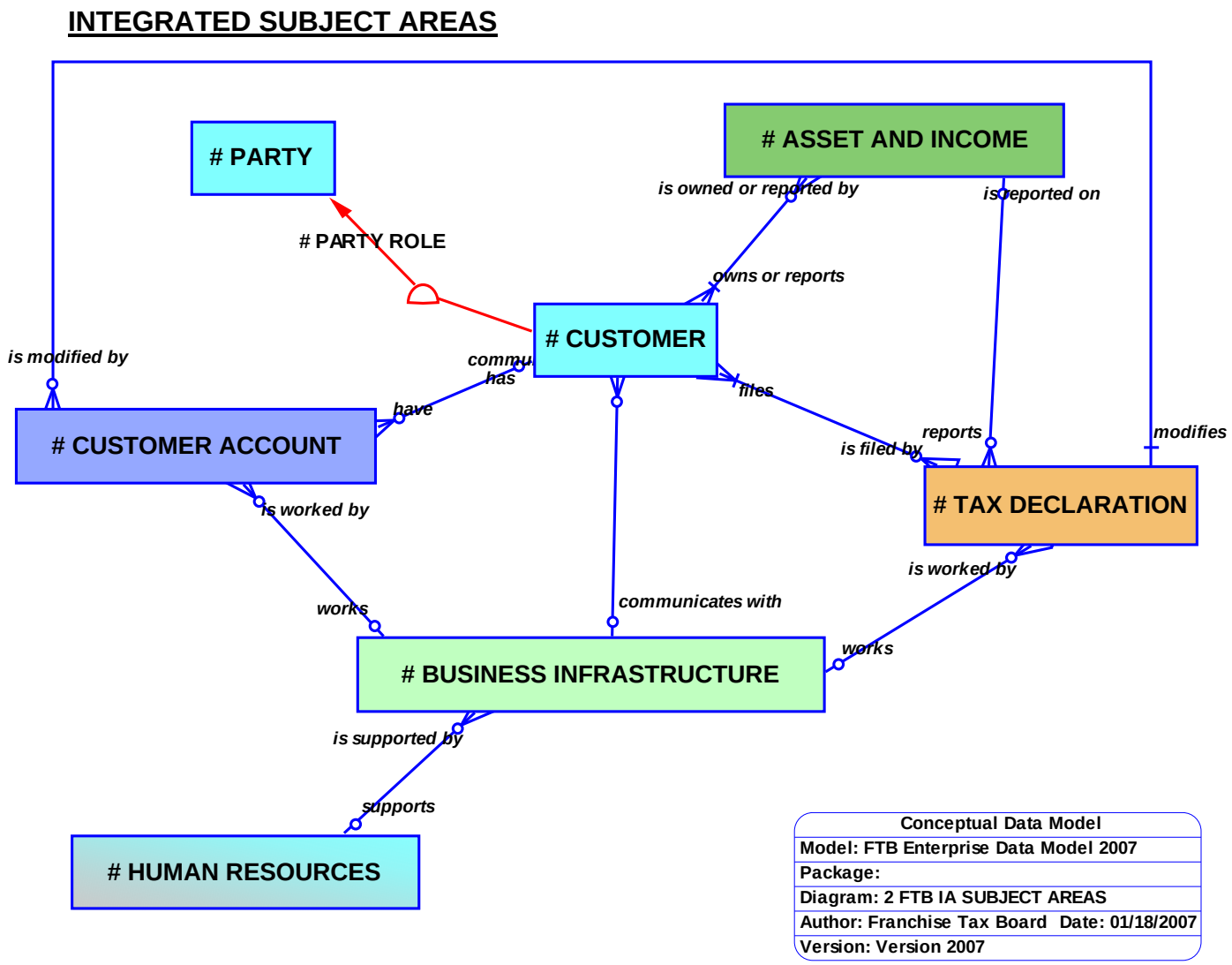
Applications are the day-to-day use of core business data. Partitioning applications allows for 1) consistent editing using enterprise business rules, and 2) production of management reports without affecting the performance of operational systems.

- **Information processing consists of the acquisition and editing (cleaning/scrubbing) of raw data.**
  Separating this function from the core business applications allows for 1) the consistent use of business rules and the establishment of enterprise pools of clean data available for multiple applications, and 2) the ability of the core business application to concentrate on the problem to be solved.
- **Management reporting is the ability to present management with the information needed to make informed business decisions.**
  Management reporting (e.g., revenue tracking) systems should be developed separately from operational systems.

6.3 FTB Enterprise Data Architecture *(Subject Areas and Data Entities)*

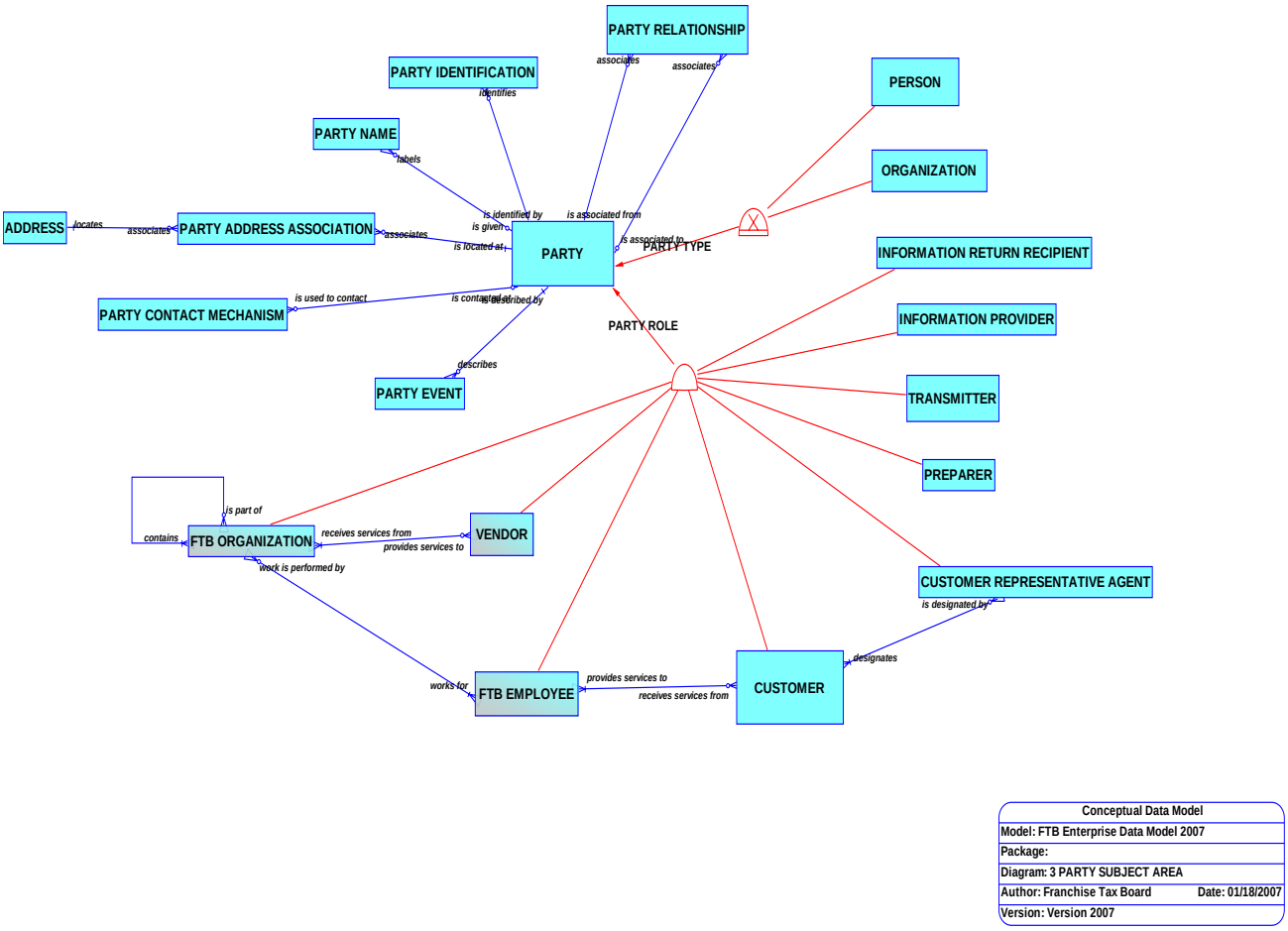
6.3.1 Subject Areas

Figure 6.3-1: Integrated Subject Areas



PARTY SUBJECT AREA

Figure 6.3-2: Party Subject Area

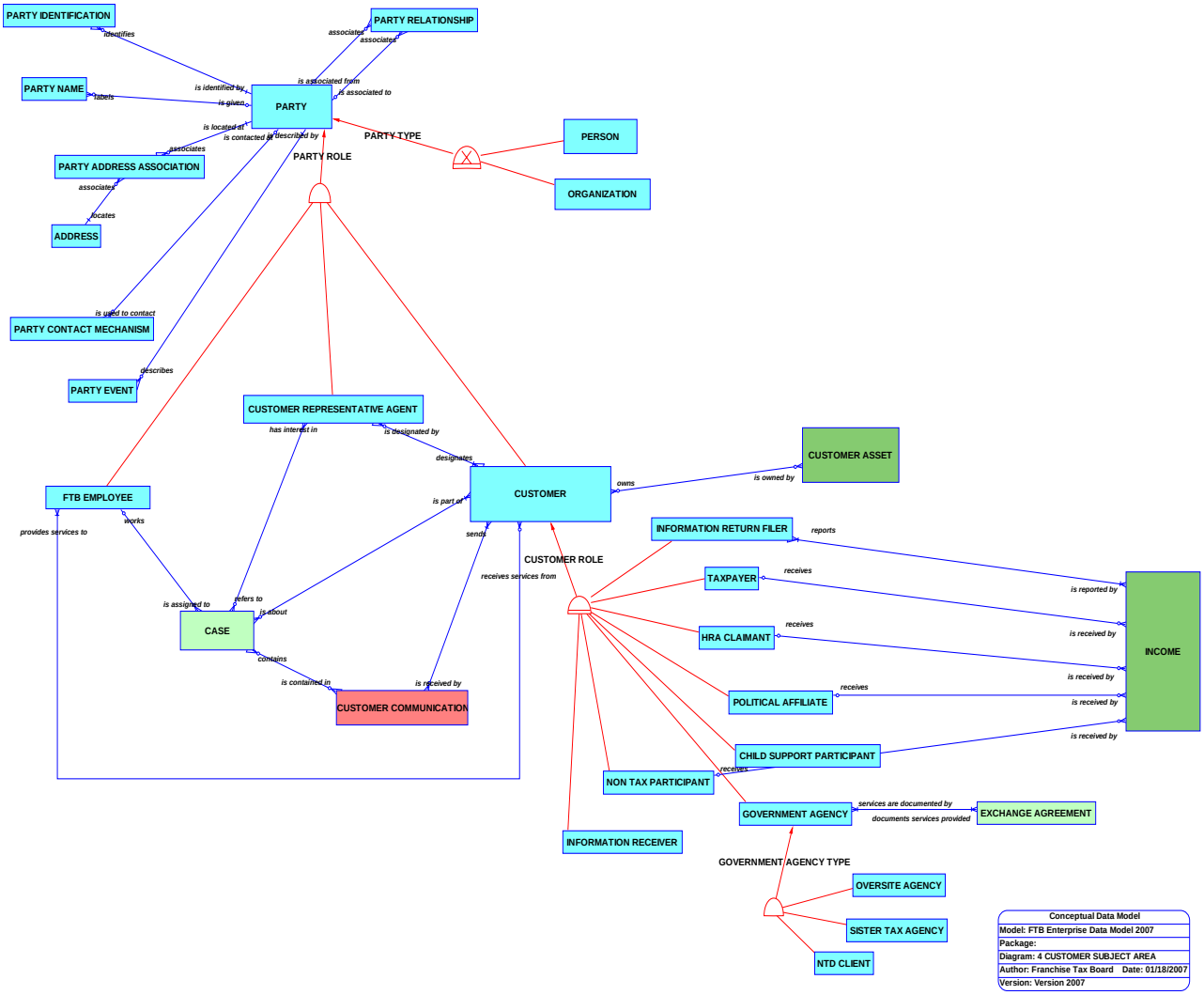


Description PARTY SUBJECT AREA

PARTY SUBJECT AREA is a grouping of data about an individual or organization that does business with the FTB.

CUSTOMER SUBJECT AREA

Figure 6.3-3: Customer Subject Area

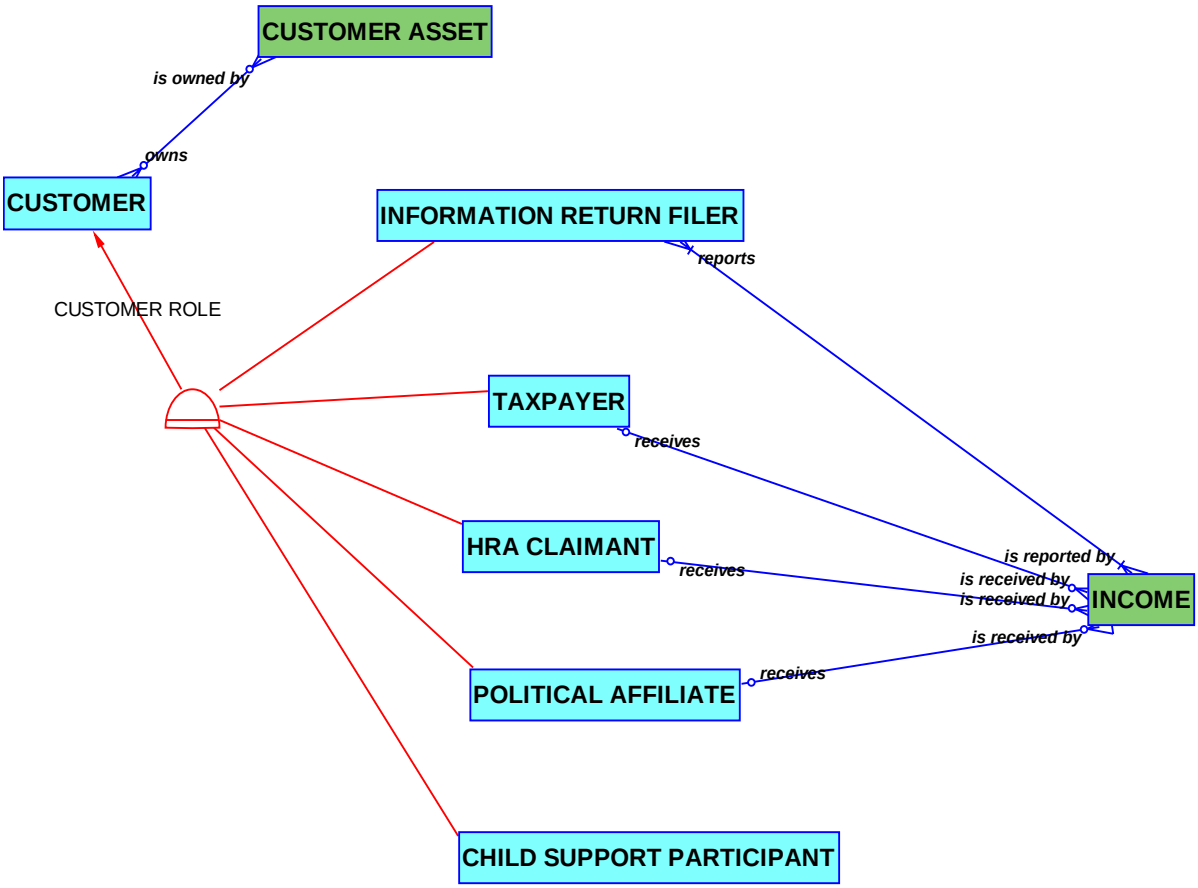


Description CUSTOMER SUBJECT AREA

CUSTOMER SUBJECT AREA is a grouping of data about an individual or business organization that does business or falls within the jurisdiction of law(s) of the State of California administered and/or enforced by the FTB.

ASSET AND INCOME SUBJECT AREA

Figure 6.3-4: Asset and Income Subject Area



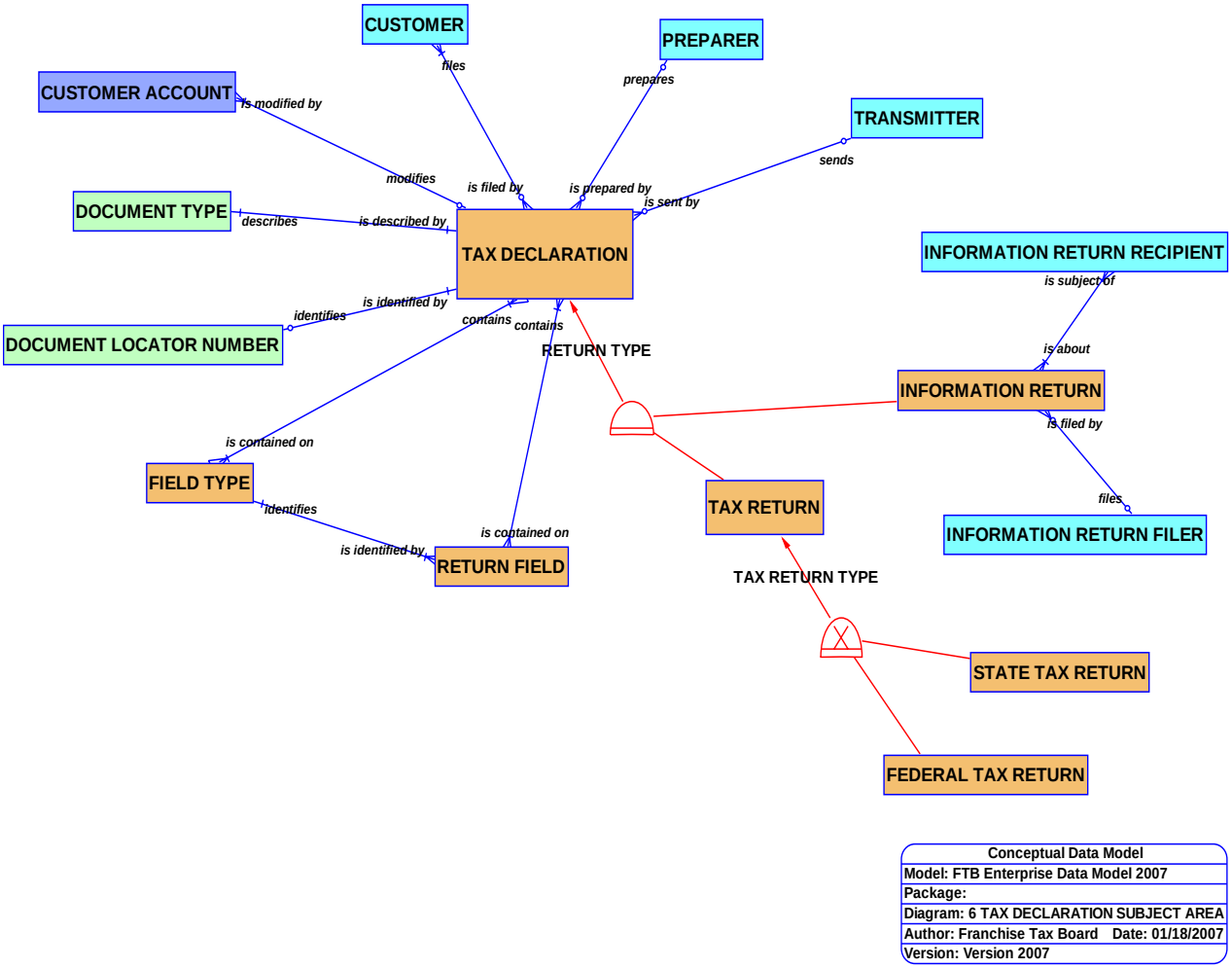
Conceptual Data Model	
Model: FTB Enterprise Data Model 2007	
Package:	
Diagram: 5 ASSET AND INCOME SUBJECT AREA	
Author: Franchise Tax Board	Date: 01/18/2007
Version: Version 2007	

Description ASSET AND INCOME SUBJECT AREA

ASSET AND INCOME SUBJECT AREA is a grouping of data about the CUSTOMER's asset and income sources and amounts, which are of interest to the FTB.

TAX DECLARATION SUBJECT AREA

Figure 6.3-5: Tax Declaration Subject Area



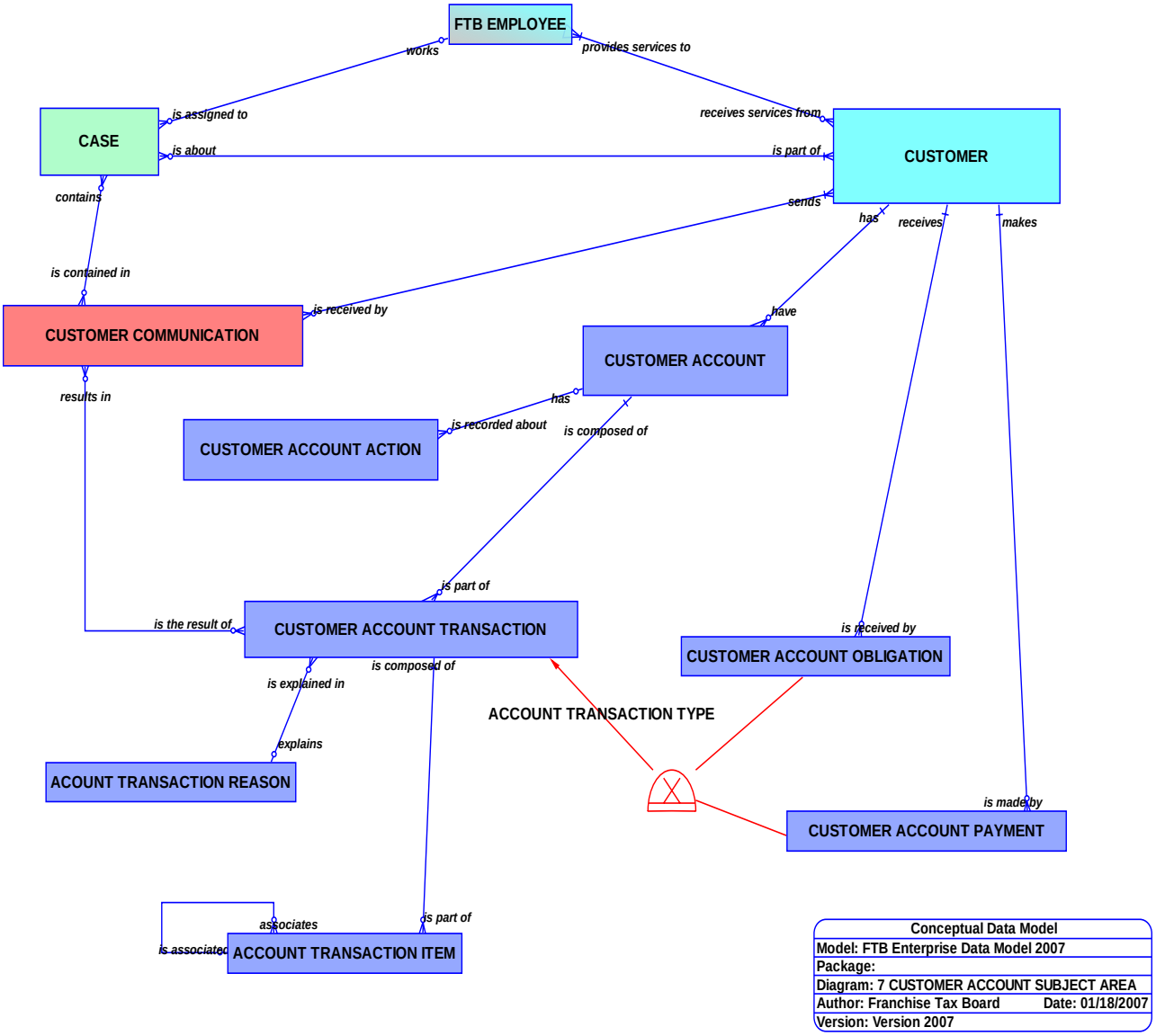
Description TAX DECLARATION SUBJECT AREA

TAX DECLARATION SUBJECT AREA is a grouping of data about official documents (electronic or paper) used by customers for the purpose of reporting information relating to financial data, tax liability, earnings, and other required information. A Tax Declaration is composed of various fields, which may contain an amount or descriptive information.

Examples: Tax Returns, e-filing, Information Returns, etc.

CUSTOMER ACCOUNT SUBJECT AREA

Figure 6.3-6: Customer Account Subject Area



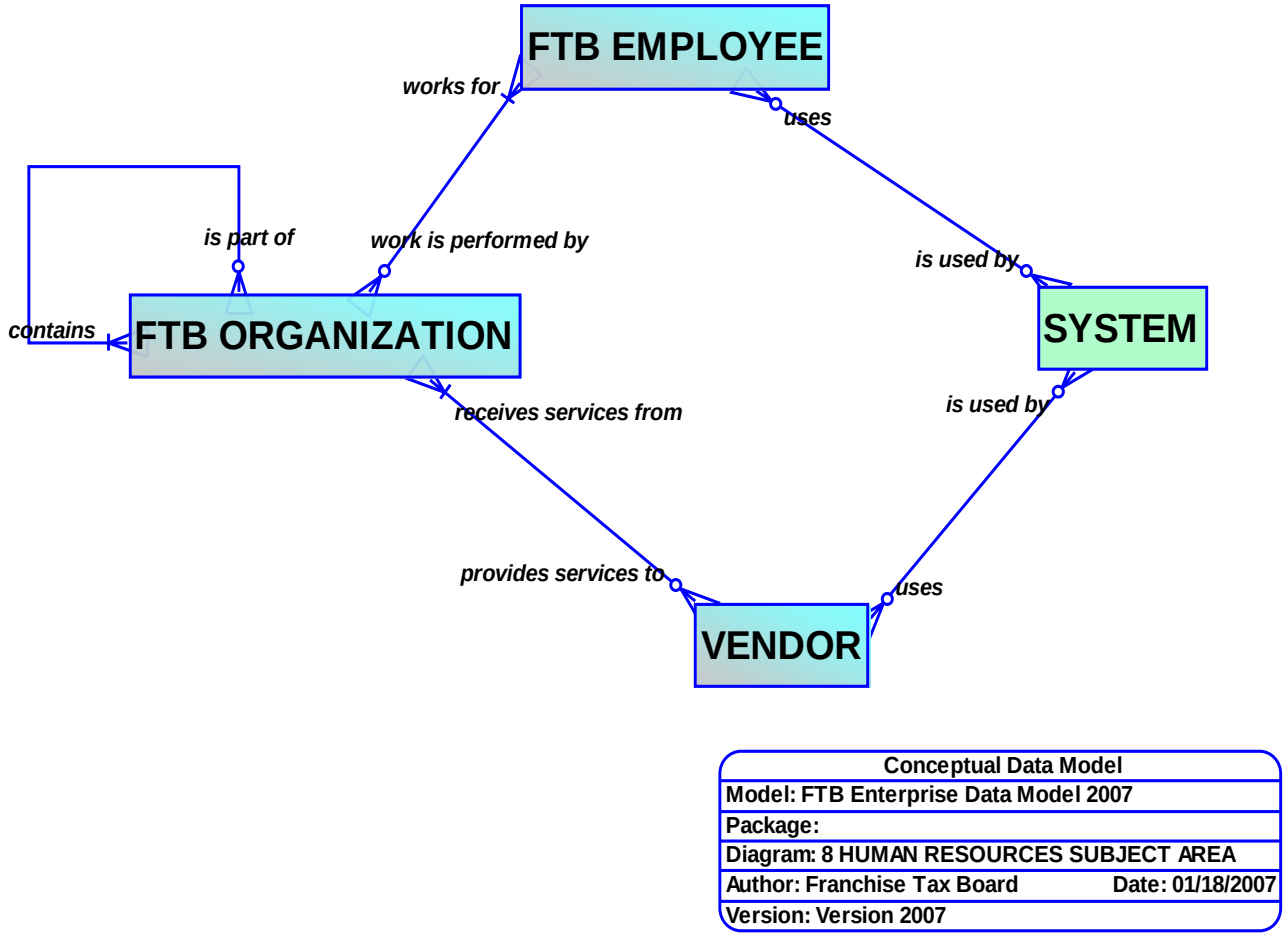
Description CUSTOMER ACCOUNT SUBJECT AREA

CUSTOMER ACCOUNT SUBJECT AREA is a grouping of transactions with the FTB. Includes payments, adjustments, obligations, liabilities, etc.

Examples: Tax Account, DMV Registration, Court-Ordered Debt, Industrial Health and Safety, etc.

HUMAN RESOURCES SUBJECT AREA

Figure 6.3-7: Human Resources Subject Area



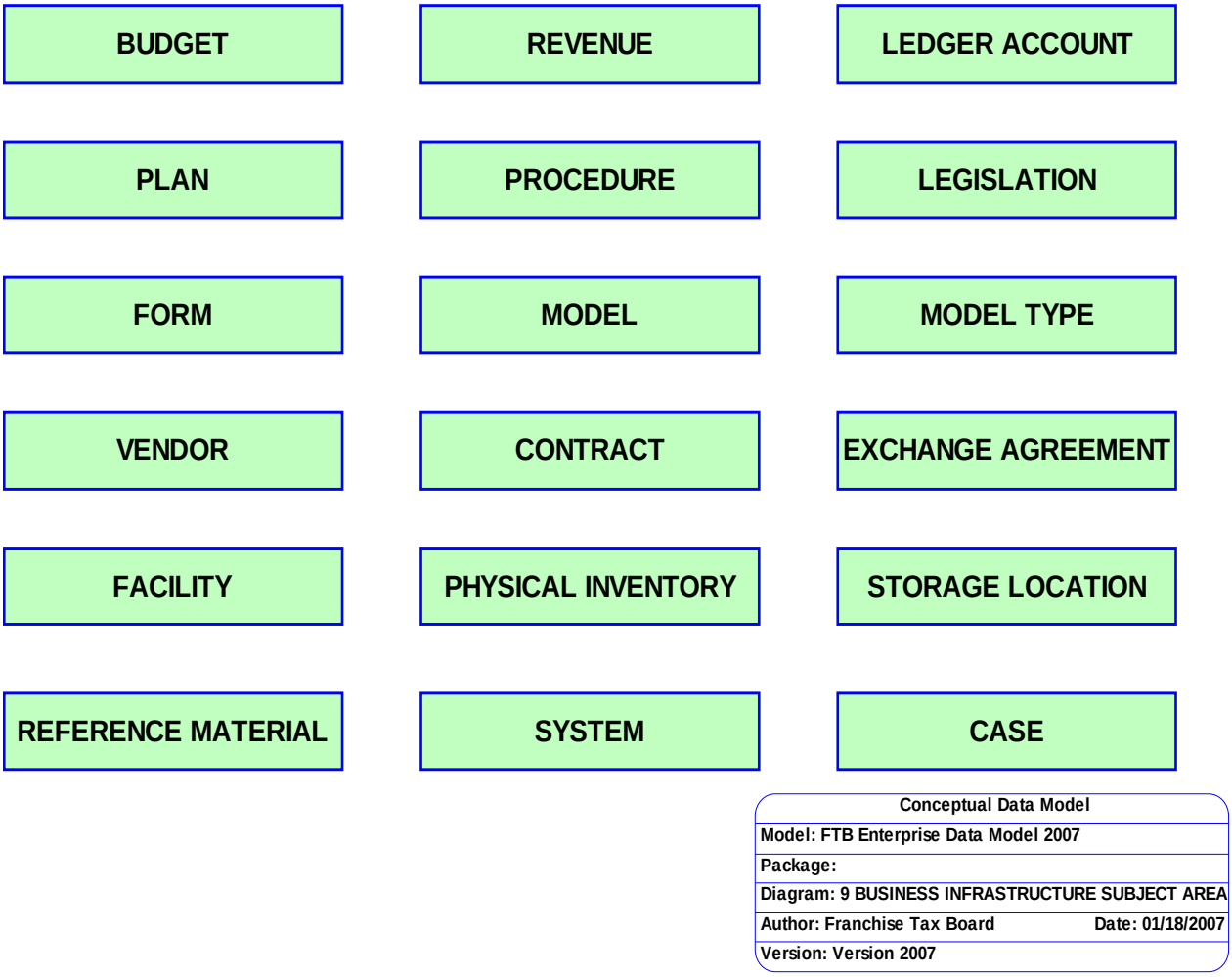
Description HUMAN RESOURCES SUBJECT AREA

HUMAN RESOURCES SUBJECT AREA is a grouping of data about the people who receive compensation directly or indirectly in the exchange for their services. This includes the characteristics, classifications, and their duties.

Examples: Employee (Auditor, Collector, Clerical, Analysts, Management, Programmers, etc.), independent contractor, vendor.

BUSINESS INFRASTRUCTURE SUBJECT AREA

Figure 6.3-8: Business Infrastructure Subject Area



Description BUSINESS INFRASTRUCTURE SUBJECT AREA

BUSINESS INFRASTRUCTURE SUBJECT AREA is a grouping of data about the functions that support the core business processes of the FTB. This includes all of the subject areas and information categories not included in the core subject areas.

Examples: Case, Models, Forms, Ledger, Accounts, Physical Inventory, Location

6.3.1 Data Entities

Name	Comment
ACCOUNT TRANSACTION ITEM	ACCOUNT TRANSACTION ITEM is data about the instance representation of a "journal entry line item" in accounting terms. According to the principles of double-entry accounting, each ACCOUNT TRANSACTION is composed of at least two ACCOUNT TRANSACTION ITEMs (i.e., a debit and a credit).
ACOUNT TRANSACTION REASON	ACCOUNT TRANSACTION REASON is data that describes for which an ACCOUNT TRANSACTION is applied.
ADDRESS	ADDRESS is data about a delivery point and/or physical location. Standardized addresses allow for the accuracy of mail delivery and save the cost of returned mail.
BUDGET	BUDGET is data about an itemized summary of probable expenditures, allocations and/or budgeted positions for a given period. Comment: Examples are training, budget, travel budget, equipment, branch budget, salaries, contracts, project budget, etc.
CASE	CASE is data about the common issues and associated activities or actions that involve persons or organization parties. Comment: Examples are Non-Tax Debt, DMV, collection, DIR, fraud, audit, protest, bankruptcy, probate, appeal, settlement, investigations, litigation, withhold at source(real estate sales, entertainment performances), etc.
CONTRACT	CONTRACT is data about an agreement between the FTB and other parties that is written and enforceable by law. Comment: Examples are postal contract, personal services contracts, equipment lease agreement, vendor contract, labor contract, etc.

Name	Comment
CUSTOMER	CUSTOMER is data about a PARTY (Person or Business Organization) that does business with or falls within the jurisdiction of law(s) of the State of California administered or enforced by the FTB. Comment: Examples are TAXPAYER, GOVERNMENT AGENCY, NON-TAX DEBTOR, INFORMATION RECEIVER, INFORMATION RETURN FILER, HRA CLAIMANT, POLITICAL AFFILIATE
CUSTOMER ACCOUNT	CUSTOMER ACCOUNT is data about the CUSTOMER's transactions with the FTB. Comment: Examples are tax, Child Support, DMV registration, court ordered debt, Industrial Health and Safety, etc. Examples of when a new customer account is created: a new return is filed, a potential non-filer is identified, assignment of non-tax collection account, filing of HRA claim, payment received for new customer, offset request for new customer, customer changes to estate or trust, etc.
CUSTOMER ACCOUNT ACTION	CUSTOMER ACCOUNT ACTION is data about the activities taken that are recorded on a CUSTOMER ACCOUNT. Comment: Examples are establish a new account, file a lien, place a hold on the account, add a comment, encumber an asset, issue an attachment, assign to collections, send a notice, EFT status (mandatory vs. voluntary), etc.
CUSTOMER ACCOUNT LIABILITY	CUSTOMER ACCOUNT LIABILITY is data about a PARTY's legal commitment to pay the FTB in a CUSTOMER ACCOUNT TRANSACTION.

Name	Comment
CUSTOMER ACCOUNT PAYMENT	CUSTOMER ACCOUNT PAYMENT is data about the remittances received or acquired by FTB that are applied to CUSTOMER ACCOUNTs. CUSTOMER ACCOUNT PAYMENTS are types of CUSTOMER ACCOUNT TRANSACTIONS. Comment: Examples are Customer Account Payment types: estimate, extension, withholding, return/remit, misc. identified, misc. unidentified, bill payment as result of RIN or STD, proposed assessment, etc. Liability Types: penalty, interest, tax, non-tax collection, fees, agency penalties, etc. Examples of Origin: asset seizures, bank, employer, withholding agent, garnishments, levies, escrows, bankruptcies, probate proceedings, etc. Media Types: EFT, paper, cash, Western Union, etc.
CUSTOMER ACCOUNT TRANSACTION	CUSTOMER ACCOUNT TRANSACTION is data about the representation of a journal entry in accounting terms. Note that ACCOUNT TRANSACTION does not have transaction amounts as this is calculated by summing the line items of the ACCOUNT TRANSACTION ITEM entity. Each ACCOUNT TRANSACTION must be composed of one or more ACCOUNT TRANSACTION ITEMs. CUSTOMER ACCOUNT TRANSACTION is data about a credit or debit that results in an addition, deletion or revision to a CUSTOMER's ACCOUNT. Comment: Examples are post a transfer to a CUSTOMER ACCOUNT, post a payment to a CUSTOMER ACCOUNT, etc.
CUSTOMER ASSET	CUSTOMER ASSET is data about anything a CUSTOMER possesses in which they have an ownership interest. Comment: Examples are 1. Real property (e.g., land, buildings), personal property (e.g., inventory, furnishings, boats, cars, stocks, bonds, cash, bank accounts, dividend income, accounts receivable), etc. 2. Monetary or financial obligation owed to the TAXPAYER by third parties including government agencies or unpaid wages (e.g., refund due a taxpayer), etc.

Name	Comment
CUSTOMER COMMUNICATION	CUSTOMER COMMUNICATION is data about any interaction between PARTY and the FTB. Comment: Examples are sending forms, requests for information, inquiry, phone contact notes, letters sent, letters received, educational material, notices, etc.
CUSTOMER REPRESENTATIVE AGENT	CUSTOMER REPRESENTATIVE AGENT is data about a PARTY (Person or Organization) who represents the interest of an FTB CUSTOMER.
DOCUMENT LOCATOR NUMBER	DOCUMENT LOCATOR NUMBER is data about the physical location of documents received and processed. Comment: DLN includes: type of RETURN, truck #, block #, process year, sequence number
DOCUMENT TYPE	DOCUMENT TYPE is data about the kind of form, schedule, statement, claim filed by a CUSTOMER. Comment: Examples are 540, 9000, 1099, etc.
ENTITY	ENTITY is data about a group of individuals (persons) or a legal organization (business organization) formed for a common purpose. Comment: Examples of ENTITY types: corporations, non-profit corporations, partnerships, associations, trusts, banks and financials, government agencies, political committee, lobbyist group, trusts, LLC, LLP, cooperatives, S corps, C corps, master limited partnerships, non-admitted insurance co., estates, RICS, REMICS, REITS, etc. Examples of ENTITY status: qualified vs. non-qualified, franchise vs. income tax.
EXCHANGE AGREEMENT	EXCHANGE AGREEMENT is data about the written documentation for the exchange of information or services between the FTB and other GOVERNMENT AGENCY(s).
FACILITY	FACILITY is data about an existing or proposed structure used to house the FTB staff, equipment, and/or materials. Comment: Examples are District Offices, Central Office, Sun Center, Trade Center, conference rooms, training rooms, security access, etc.

Name	Comment
FEDERAL TAX RETURN	FEDERAL TAX RETURN is data about the official documents used by a TAXPAYER to report their demographics, earnings, financial data, tax liability and other required information to the Internal Revenue Service.
FIELD TYPE	FIELD TYPE is data about the identification of different line items or monetary information reported on RETURNS. Comment: Examples are mortgage interest, interest paid, dividend, tax liability, refund, payment, credit, income, deduction, etc.
FORM	FORM is data about an approved, numbered document used to obtain and/or provide information. This includes electronic forms. Comment: Examples are tax returns/schedule, information returns, leave requests, bills, account action requests, timesheets (634), Interagency Agreement, etc.
FTB EMPLOYEE	FTB EMPLOYEE is data about PERSON (PARTY) who works or has worked for the FTB and is subject to civil service rules. Does not include independent contract workers.
FTB ORGANIZATION	FTB ORGANIZATION is data about a PARTY that is a Branch, Division, Bureau, Section, Unit or other group within the FTB.
GOVERNMENT AGENCY	GOVERNMENT AGENCY is data about a CUSTOMER that is a public organization with whom the FTB does business.
HRA CLAIMANT	HRA CLAIMANT is data about a CUSTOMER who is a senior citizen (62 years of age or older), blind, or disabled individual that may be eligible for property tax relief or rent assistance.
INCOME	INCOME is data about the amount of money or its equivalent received by a CUSTOMER from any source (taxable or nontaxable).
INFORMATION PROVIDER	INFORMATION PROVIDER is data about PARTYs (Person or Organization) that provide information to the FTB other than through EXCHANGE AGREEMENTS or required by information reporting laws.
INFORMATION RECEIVER	INFORMATION RECEIVER is data about a CUSTOMER (Person or Business Organization) who receives appropriate information from the FTB in accordance with disclosure laws.

Name	Comment
INFORMATION RETURN	INFORMATION RETURN is data about the official documents used for the purpose of reporting payments and/or information required by information reporting laws.
INFORMATION RETURN FILER	INFORMATION RETURN FILER is data about a CUSTOMER who is subject to information reporting laws.
INFORMATION RETURN RECIPIENT	INFORMATION RETURN RECIPIENT is data about a PARTY (Person or Organization) that is the subject of an INFORMATION RETURN.
LEDGER ACCOUNT	LEDGER ACCOUNT is data about the ledger maintained for FTB ORGANIZATION(s) and their support appropriations. Comment: Examples are budget plans, revenue accounts, appropriations, fiscal account, receivables, etc.
LEGISLATION	LEGISLATION is data about a proposed or enacted law or group of laws relevant to the FTB. Comment: Examples are wage reporting simplification (AB3086), Revenue and Taxation Code, Internal Revenue Code, Taxpayer Bill of Rights, Pension Act, etc.
MODEL	MODEL is data about the particular set of selection criteria used to develop a workload or identify trends. Comment: Examples are Solar Energy Credit, Head of Household with Bad Address, Research and Statistics Model, etc.
MODEL TYPE	MODEL TYPE is data about a specific category of MODEL. Comment: Examples are Audit Model, Collection Model, Research and Statistics Model, etc.
NON TAX PARTICIPANT	NON-TAX PARTICIPANT is data about a CUSTOMER (Person or Business Organization) that has a non-tax liability, which the FTB is responsible for collecting.
NTD CLIENT	NTD CLIENT is data about a party that is a public organization with whom the FTB does business for non-tax debt. Client Agencies are: Federal, State, County, Court, Office, City
ORGANIZATION	ORGANIZATION is data about a group or business that does business with FTB.

Name	Comment
OVERSITE AGENCY	OVERSIGHT AGENCY is data about a GOVERNMENT AGENCY - CUSTOMER that oversees FTB activities.
PARTY	PARTY is data about a person or organization that has a role in doing business with the FTB.
PARTY ADDRESS ASSOCIATION	PARTY ADDRESS ASSOCIATION is data about the associations between PARTIES and Addresses. The PARTY ADDRESS ASSOCIATION allows for the standardization of addresses, which reduces redundancy and improves quality in the location of PARTIES. An example of this association allows a historical tracking of PARTY addresses, the types of address, which PARTIES are associated to one address, and which ADDRESSes are associated to one PARTY.
PARTY CHARACTERISTIC	PARTY CHARACTERISTIC is data about the behavior patterns of a party. Examples: potentially dangerous, sensitive, bad check writer, etc
PARTY CONTACT MECHANISM	PARTY CONTACT MECHANISM is data about the way to contact a PARTY and the location of that contact mechanism. Examples of a PARTY CONTACT MECHANISM are: Phone Numbers, Fax Numbers, Pager Numbers, Cellular Numbers, e-mail addresses, etc.
PARTY EVENT	PARTY EVENT is data about the condition and when it occurred of the PARTY. Comment: Examples are deceased, bankrupt, dissolved, suspended, etc.
PARTY IDENTIFICATION	CUSTOMER IDENTIFICATION is data about the ways the FTB identifies CUSTOMERs. Comment: Examples are TPID, SSN, FEIN, Corporate Account Number, Drivers License Number, BEAN, SEIN (EDD Account Number), etc.
PARTY NAME	PARTY NAME is data about the name that is given to a person or an organization.
PARTY RELATIONSHIP	PARTY RELATIONSHIP is data about the connections between PARTIES who have a joint or combined reporting requirement.
PERSON	A PERSON is data about a PARTY that is an individual human being that does business or interacts with FTB.

Name	Comment
PHYSICAL INVENTORY	PHYSICAL INVENTORY is data about supplies and equipment purchased or leased. Comment: Examples are desks, chairs, typewriters, pencils, copy machines, PC's, etc.
PLAN	PLAN is data about a coordinated set of tasks, visions, goals, objectives, drawings, and/or models to achieve a specific purpose. Comment: Examples are strategic plan, project plan, operational plan, tactical plan, etc.
POLITICAL AFFILIATE	POLITICAL AFFILIATE is data about CUSTOMER who is a candidate, lobbyist, political action committee, or ballot measure campaign in the State of California. This is of interest to comply with the Political Reform Act of 1974.
PREPARER	PREPARER is data about a PARTY (Person or Organization) compensated for completing a return.
PROCEDURE	PROCEDURE is data about a prescribed course of action. Comment: Examples are audit procedures, collection procedures, cashiering procedures, desk manuals, Employee Nomination Award, security procedures, etc.
REFERENCE MATERIAL	REFERENCE MATERIAL is data about the internal and external information sources used by FTB employees to perform the functions of their jobs. Comment: Examples are law books, manuals, Lexis/Nexis, internet/intranet, legislation, State Administration Manual, reverse directories, phone books, property tax records, departmental policy, case law, etc.
RETURN FIELD	RETURN FIELD is data recorded in a specific area of a RETURN. It may be either an amount or descriptive information. Comment: Examples are on a Form 540, Exemption field may contain a number; on a Form 540, Name field may contain the taxpayer name; on a Form 1099, Rent field may contain a monetary amount, etc.
REVENUE	REVENUE is data about monies assessed or collected by the FTB. Comment: Examples are dollars assessed from tax returns and audit assessments, dollars collected from tax payments, involuntary collections (attachments), and contract work, etc.

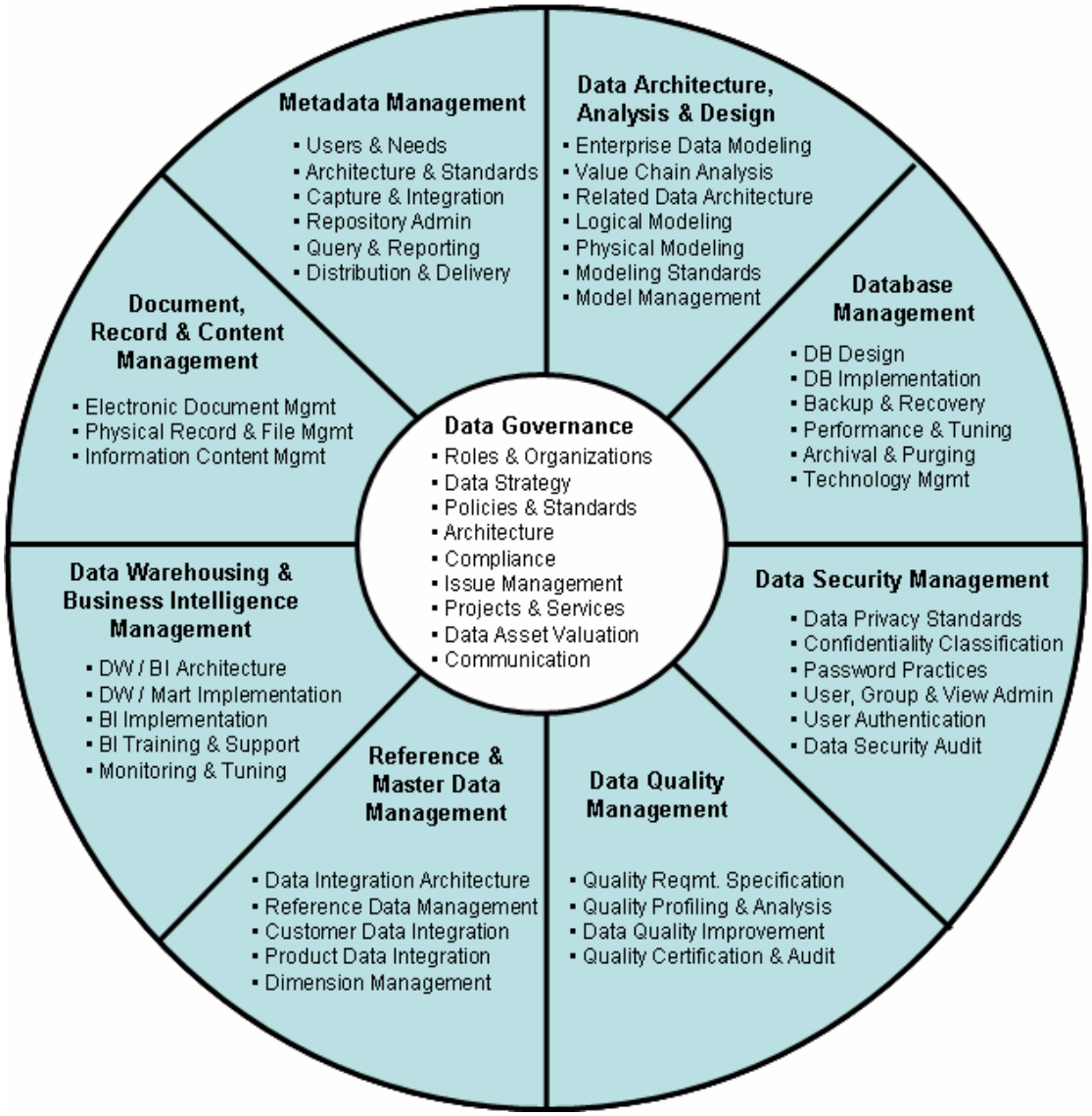
Name	Comment
SISTER TAX AGENCY	SISTER TAX AGENCY is data about a GOVERNMENT AGENCY - CUSTOMER with which the FTB has similar business interests.
STATE TAX RETURN	TAX RETURN is data about the official documents used by a TAXPAYER to report their demographics, earnings, financial data, tax liability and other required information.
STORAGE LOCATION	STORAGE LOCATION is data about a space for storing items of interest to the FTB. Comment: Examples are forms warehouse, Information Storage, supply room, trucks, file cabinets, PCs, tape library, vault, law library, etc.
SYSTEM	SYSTEM is data about the integration of data, software, hardware, methods, and/or network that process and provide information to the FTB.
TAX DECLARATION	TAX DECLARATION is data about official documents (whether paper of electronic) used by CUSTOMERs for the purpose of reporting information relating to financial data, tax liability, earnings, and other required information. A return is composed of various fields, which may contain an amount or descriptive information.
TAX RETURN	TAX RETURN is data about the official documents used by a TAXPAYER to report their demographics, earnings, financial data, tax liability, and other required information.
TAXPAYER	TAXPAYER is data about a CUSTOMER (Person or Business Organization) that falls within the jurisdiction of the Revenue and Taxation Code of the State of California administered or enforced by the FTB.
TRANSMITTER	TRANSMITTER is data about a PARTY (Person or Organization) who transmits RETURNs and/or CUSTOMER ACCOUNT PAYMENTs and may be compensated.
VENDOR	VENDOR is data about an external PARTY (person or business organization) that provides services, materials and/or equipment needed by the FTB.

6.4 Industry Best Practices

6.4.1 Data Management Body of Knowledge (DMBOK).

The diagram below illustrates the nine functions within the DMBOK framework.

Figure 6.4-1: Data Management Associations (DAMA) Data Management Framework



DMBOK further breaks down each function into sub-functions and activities. Each activity has been categorized as belonging to one of four Activity Groups. FTB can choose to use these activity assignments when making organizational responsibility assignments.

- Planning Activities (P) — Activities that set the strategic and tactical course for other data management activities. Planning activities may be performed on a recurring basis.
- Control Activities (C) — Supervisory activities performed on an on-going basis.
- Development Activities (D) — Activities undertaken within projects and recognized as recognized as part of the systems development lifecycle (SDLC), creating data deliverables through analysis, design building, testing and deployment.
- Operational Activities (O) — Service and support activities performed on an on-going basis.

1. Data Governance

1.1. Data Management Planning

1.1.1. Identify Strategic Enterprise Data Needs (P)
1.1.2. Develop and Maintain the Data Strategy (P)
1.1.3. Establish the Data Management Professional Organizations (P)
1.1.4. Identify and Appoint Data Stewards (P)
1.1.5. Establish Data Governance and Stewardship Organizations (P)
1.1.6. Develop, Review and Approve Data Policies, Standards and Procedures (P)
1.1.7. Review and Approve Data Architecture (P)
1.1.8. Plan and Sponsor Data Management Projects and Services (P)
1.1.9. Estimate Data Asset Value and Associated Costs (P)

1.2. Data Management Supervision and Control

1.2.1. Supervise the Data Management Professional Staff and Organizations (C)
1.2.2. Coordinate Data Governance Activities (C)
1.2.3. Manage and Resolve Data Related Issues (C)
1.2.4. Monitor and Ensure Regulatory Compliance (C)
1.2.5. Communicate, Monitor and Enforce Conformance with Data Policies, Standards, Procedures and Architecture (C)
1.2.6. Oversee Data Management Projects and Services (C)
1.2.7. Communicate and Promote the Value of Data Assets (C)

2. Data Architecture, Analysis and Design

2.1. Enterprise Data Architecture

2.1.1. Develop the Enterprise Data Model (P)
2.1.2. Align with Other Business Models (P)
2.1.3. Define the Database Architecture (P) (same as 3.3.2)
2.1.4. Define the Data Integration / MDM Architecture (P) (same as 6.2)
2.1.5. Define the Data Warehouse / BI Architecture (P) (same as 7.2)
2.1.6. Define the Metadata Architecture (P) (same as 9.2)
2.1.7. Define Enterprise Taxonomies and Namespaces (P)

2.2. Data Modeling and Specification

2.2.1. Define Information Needs (D)
2.2.2. Develop and Maintain Logical Data Models (D)
2.2.3. Develop and Maintain Physical Data Models (D)

2.3. Data Model Quality Management

2.3.1. Develop Data Modeling Standards (P)

2.3.2. Review Data Model Quality (C)
2.3.3. Manage Data Model Versioning and Integration (C)

3. Database Management

3.1. Database Development

3.1.1. Define Database Design Standards (P)
3.1.2. Design Physical Databases (D)
3.1.3. Review Database Design Quality (C)
3.1.4. Develop Data Access Services (D)
3.1.5. Develop Information Products ((D)
3.1.6. Implement Development / Test Database Changes (C)
3.1.7. Create and Maintain Test Data (D)
3.1.8. Migrate and Convert Data (D)
3.1.9. Test and Validate Data Requirements (D)

3.2. Database Production Support

3.2.1. Implement Production Database Changes (C)
3.2.2. Obtain Externally Sourced Data (O)
3.2.3. Plan for Data Recovery (P)
3.2.4. Backup and Recover Data (O)
3.2.5. Set Database Performance Service Levels (P)
3.2.6. Monitor and Tune Database Performance (O)
3.2.7. Plan for Data Retention (P)
3.2.8. Archive, Retrieve and Purge Data (O)
3.2.9. Manage Specialized Databases (O)

3.3. Data Technology Management

3.3.1. Understand Data Technology Requirements (P)
3.3.2. Define the Database Architecture (P) (same as 2.1.3)
3.3.3. Implement and Maintain Database Environments (C)
3.3.4. Evaluate Data Technology (P)
3.3.5. Install and Administer Data Technology (O)
3.3.6. Inventory and Track Data Technology Licenses (C)
3.3.7. Support Data Technology Usage and Issues (O)

4. Data Security Management

4.1. Understand Data Privacy, Confidentiality and Security Needs (P)
4.2. Define Data Privacy and Confidentiality Standards (P)
4.3. Define Password Standards and Procedures (P)
4.4. Implement Data Security Controls (D)
4.5. Manage Users, Passwords and Group Membership (C)
4.6. Manage Data Access Views (C)
4.7. Manage Data Access Permissions (C)
4.8. Monitor User Authentication and Access Behavior (C)
4.9. Classify Information Confidentiality (C)
4.10. Audit Data Security (C)

5. Data Quality Management

5.1. Develop and Promote Data Quality Awareness (O)
5.2. Define Data Quality Metrics (P)
5.3. Define Data Quality Requirements and Business Rules (D)
5.4. Analyze / Profile / Measure / Monitor Data Quality (C)
5.5. Set Data Quality Service Levels (P)

5.6. Certify Data Quality (C)
5.7. Identify, Escalate and Resolve Data Quality Issues (C)
5.8. Conduct Clean-up Campaigns (O)
5.9. Design and Implement Operational DQM Procedures (D)
5.10. Monitor Operational DQM Procedures (C)
5.11. Test and Validate Data Quality Requirements (D)
5.12. Audit Data Quality (C)

6. Reference and Master Data Management

6.1. Understand Reference and Master Data Integration Needs (P)
6.2. Define the Data Integration / MDM Architecture (P) (same as 2.1.4)
6.3. Implement Reference and Master Data Management Solutions (D)
6.4. Control Code Values and Other Reference Data (C)
6.5. Integrate Master Data (O)
6.6. Replicate Reference and Master Data (O)
6.7. Maintain Dimensional Hierarchies (O)

7. Data Warehousing and Business Intelligence Management (See EA Definition BI V1.doc)

7.1. Understand Business Intelligence Data Needs (P)
7.2. Define the Data Warehouse / BI Architecture (P) (same as 2.1.5)
7.3. Implement Data Warehouses and Data Marts (D)
7.4. Implement Business Intelligence Tools and User Interfaces (D)
7.5. Implement Enterprise Reporting (D)
7.6. Implement Management Dashboards and Scorecards (D)
7.7. Implement Analytic Applications (D)
7.8. Train Business Professionals (O)
7.9. Replicate and Transform Data for Business Intelligence (O)
7.10. Monitor and Tune Data Warehousing Processes (C)
7.11. Support Business Intelligence Activity (O)
7.12. Monitor and Tune BI Activity and Performance (C)

These activities do not include actual Business Intelligence activities performed by knowledge workers, which include:

- Perform Ad Hoc Query and Reporting
- Perform Multidimensional Analysis
- Perform Statistical Analysis
- Perform Data Mining
- Model “What If” Scenarios
- Monitor and Analyze Business Performance

8. Document, Record and Content Management

8.1. Manage Electronic Documents (text, graphics, image, audio, video)
8.2. Manage Physical Records (paper, fiche)
8.3. Manage Information Content (search engine indexes, taxonomies, XML namespaces, report and document format standards)

9. Meta Data Management

9.1. Understand Meta Data Requirements (P)
9.2. Define the Meta Data Architecture (P) (same as 2.1.6)
9.3. Develop and Maintain Meta Data Standards (P)
9.4. Implement a Managed Meta Data Environment (D)

9.5. Create, Capture, Store and Maintain Meta Data (O)
9.6. Maintain Meta Data Source Data Stores
9.7. Extract, Reconcile, Integrate and Share Meta Data (C)
9.8. Manage the Meta Data Repository (C)
9.9. Query, Report and Analyze Metadata (O)
9.10. Manage Meta Data Distribution and Delivery to Glossaries, Directories and Other Meta Data Marts (C)

6.5 Industry Trends

6.5.1 Database Federation

Trends to watch include improvements in database federation, which is the ability to provide data from a single request where data is housed in two disparate systems. In the future, these disparate systems could reside outside the department. Other departments will also have the need to access the FTB data in this fashion. These solutions will need to maintain a proper level of security. Single requests for data could also include a combination of structured and unstructured data.

6.5.2 Operational BI

Another trend is the growing use of operational data warehousing which combines traditional operational relational database with data warehouse to support real time analytics, data warehouse appliances, and column-vectored or vertical database management systems. (Also see BI Architecture Definition Document)

6.5.3 XML and Data Services

XML is becoming the standard by data delivery services to share data across systems as well as internal web services. XML appliances are a trend to address the current performance issues sometimes associated with XML. (Also see Service Oriented Architecture document)