DATA INTEGRATION GLOSSARY

U.S. Department
of Transportation

Data Integration Glossary

OFFICE OF
ASSET
MANAGEMENT

U.S. Department
of Transportation

NOTE FROM THE DIRECTOR

Office of Asset Management, Infrastructure Core Business Unit, Federal Highway Administration

This glossary is one of a series of documents on data integration being published by the Federal Highway Administration's Office of Asset Management. It defines in simple and understandable language a broad set of the terminologies used in information management, particularly in regard to database management and integration. Our objective is to provide convenient reference material that can be used by individuals involved in data integration activities. The glossary is limited to the more fundamental concepts and taxonomies applied in transportation database management and is not intended to be comprehensive.

The importance of data integration in implementing Asset Management processes cannot be overstated. My office will continue to provide information and support to all transportation agencies as they work to integrate their data.

Madeleine Bloom

Madeleine Bloom

Director, Office of Asset Management

LIST OF TERMS

aggregate data

Data that have been combined in collective or summary form (see also *atomic data*).

application

Short for application program, which is a software program designed to perform a specific function directly for the user or, in some cases, for another software program. Examples of applications include database programs, word processors, and graphics tools.

application integration

The process of bringing data or functions from one application together with those of another application using real-time or near real-time communication. Integration involves the ability to use interfaces employed by different applications.

application program interface (API)

The specific method specified by a computer operating system or program through which a programmer can make requests to the operating system of another application. An API can be differentiated with a graphical user interface (GUI) or a command interface, which are direct user interfaces to an operating system or a program (see also *graphical user interface*).

archive

A collection of computer files that have been packaged together for backup, for transfer to some other location, for saving away from the computer so that more hard disk storage can be made available, or for some other purpose. An archive can include a simple list of files or files organized under a directory or catalog structure, depending upon how a particular program supports archiving.

Asset Management

A strategic approach to managing transportation infra-structure. Asset Management requires integrated data and information to make comprehensive decisions regarding assets.

atomic data

Data items containing the lowest level of detail. For example, in a daily maintenance activity report, the individual equipment used would be atomic data, while rollups like summary totals from equipment rental invoices are *aggregate data* (see also aggregate data).

authorization request

A request initiated by a user to access a database for which he or she does not have access privileges. The criteria used to evaluate this request are called the "authorization rule."

bulk data transfer

A computer-based procedure designed to move large data files. The procedure usually involves data compression, blocking, or buffering to maximize data transfer rates.

business process

A group of activities that takes one or more types of inputs to produce an output that is of value to customers. For example, pavement repair is a business process consisting of several tasks using agency labor, materials, and equipment to restore the smoothness of the roadway.

business process reengineering (BPR)

The reexamination and redesign of business processes with the aim of achieving significant improvements in system performance measures such as cost, quality, safety, and reliability.

communications protocol

A set of conventions that governs the communications between processes. These conventions specify the format and content of messages to be exchanged and allow different computers using different software to communicate.

Computer Aided Software Engineering (CASE) tools

A class of software that provides a controlled development environment for computer programming teams. CASE systems offer tools to automate, manage, and simplify the program development process. These tools can include software for summarizing initial requirements, developing data flow diagrams, scheduling development tasks, preparing documentation, controlling software versions, and developing program code. While many CASE systems provide special support for object-oriented programming (see *object-oriented programming*), the term CASE can apply to any type of software development environment.

computer network

A group of two or more computer systems linked together for the purpose of communications or application distribution. There are many types of computer networks, including *local area networks (LANs)*—made up of computers that are geographically close, and *wide area networks (WANs)*—where computers are farther apart and are connected by telephone lines or radio waves (see definitions below). Most State highway agencies use both LANs and WANs to link

computers in their headquarters, divisions or districts, and remote area offices. Computers on a network are sometimes called "nodes." Computers and devices that allocate resources for a network are called "servers."

local area network (LAN)

A computer network that spans a relatively small area. Most LANs are confined to a single building or group of buildings, and connect workstations and personal computers. Each individual computer or node in a LAN has its own central processing unit with which it executes programs, but it is also able to access shared data and devices anywhere on the network. Users can also use the LAN for electronic mail. There are many different types of LANs; ethernet is the most common for personal computers. LANs are capable of transmitting data at very fast rates, much faster than data can be transmitted over a telephone line; but the distances are limited, and there is also a limit on the number of computers that can be attached to a single LAN.

wide area network (WAN)

A computer network that spans a relatively large geographical area. Typically, a WAN consists of two or more LANs. Computers connected to a WAN are often connected through public networks such as the telephone system. They can also be connected through leased lines or satellites. The largest WAN in existence is the Internet.
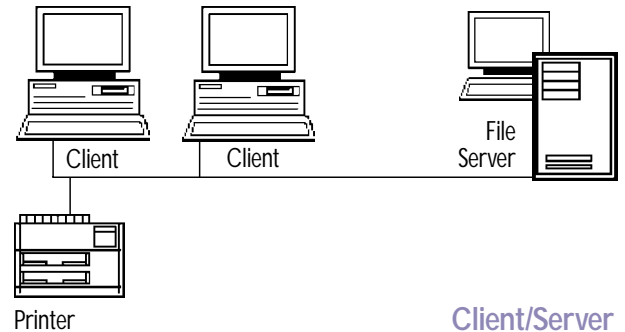
computer network architecture

Structural design that defines a computer network's general characteristics as well as its precise mechanisms. In broad terms, a computer network can have open or closed architecture. Open architectures allow the system to be connected easily to devices and programs made by other manufacturers. Open architectures use off-the-shelf components and conform to approved standards (see *open systems environment*). A system with a closed architecture, on the other hand, is one whose design is proprietary, making it difficult to connect the system to other systems. For both open and closed architectures, the client/server or peer-to-peer architecture defines the precise mechanisms by which computer network elements are tied together, as described below:

client/server architecture

A computer network architecture in which each computer or process on the network is either a client or a server. "Servers" are powerful computers or processes dedicated to managing disk drives (file servers), printers (print servers), or network traffic (network servers). "Clients" are personal computers or workstations on

which users run applications. Clients rely on servers for resources such as files, devices, and even processing power.

Client/Server

peer-to-peer architecture

A type of network architecture in which each node has equivalent responsibilities rather than being exclusively a client or a server.

Peer-to-Peer
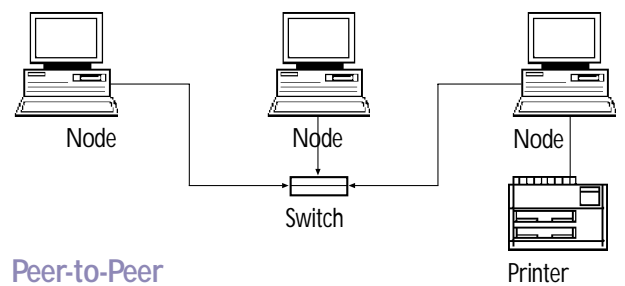
Data Exchange Format (DXF)

A proprietary but published two-dimensional graphics file format supported by virtually all PC-based computer-aided design (CAD) products. It is now a de facto standard for exchanging graphics data.

data extraction

The process of reading one or more sources of data and creating a new representation of the data.

data integration

The process of combining or linking two or more data sets from various sources to facilitate data sharing, promote effective data gathering and communication, and support overall information management activities in an organization (see also *data warehouse* and *interoperable database*).

data integrity

The degree or means by which the data in a database conform to specified data standards. For example, to maintain integrity, the numeric fields in a database will not accept alphabetic data.

data mapping

The process of assigning a source data element to a target data element.

data migration/transformation

The process of converting data from one format to another. Data migration is necessary when an organization decides to use a new computing system or database management system that is incompatible with the current system. Typically, data migration is performed by a set of customized programs or scripts that automatically transfer the data. Data migration also sometimes refers to the process of moving data from one storage device to another.

data mining

The process of extracting previously unknown, valid, and actionable information or relationships from large databases and using that information to make important decisions.

data modeling

A method used to define and analyze data requirements needed to support the business processes of an organization. The data requirements are recorded as a conceptual data model with associated data definitions. Actual implementation of the conceptual model is called a logical data model. To implement one conceptual data model may require multiple logical data models. Data modeling defines the relationships between data elements and structures.

data partitioning

The process of physically or logically dividing data into segments so it can be easily maintained or accessed. Existing relational database management systems provide this kind of distribution functionality.

data/process flow diagram

A diagram that shows the types of data produced by one business process and used as input by another. It also refers to a representation of the flow of data into, out of, and between procedures, systems, or subsystems. Data flow diagrams show the actual flow of data between designed business activities

and data stores, thus depicting how two activities interact. Process flow diagrams show the dependencies that exist for business processes but do not indicate how the dependencies may be manifested (see also *data structure diagram* and *entity relationship diagram*).

Data/Process Flow Diagram
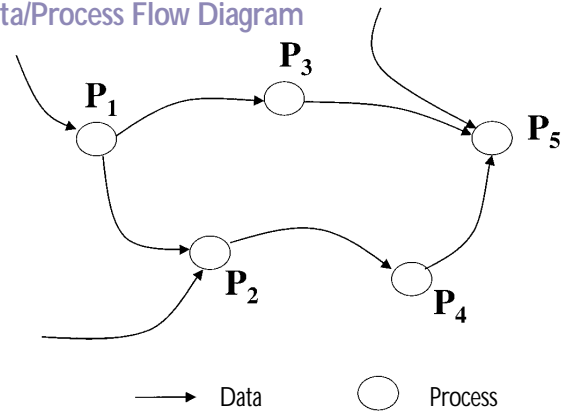
data scrubbing

The process of filtering, merging, decoding, and translating source data to create validated data for the *data warehouse* (see definition).

data structure diagram

A diagram type that is used to depict the structure of data elements in the data dictionary. The data structure diagram is a graphical alternative to the composition specifications within such data dictionary entries.

(shows composition of data elements)
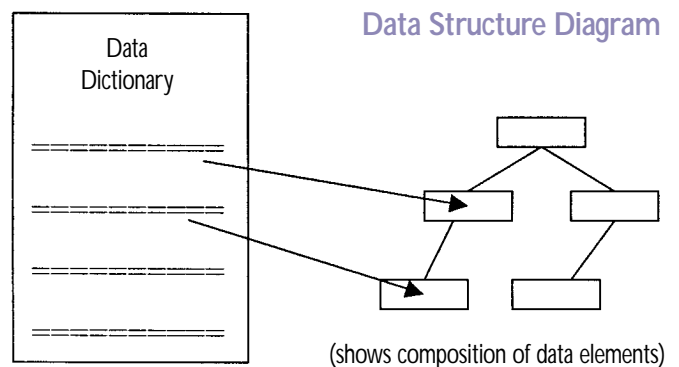
data warehouse

A collection of databases designed to support management decision-making. The term "data warehousing" generally refers to combining a wide variety of databases across an entire organization. Development of a data warehouse includes development of systems to extract data from operating systems plus installation of a warehouse database system that provides users flexible access to the data.

database (electronic)

A repository for information or data organized in such a way that a computer program can quickly select desired pieces of data. Databases are often indexed so they may be searched more efficiently. Traditional databases are organized by fields, records, and files. A field is a single piece of information; a record is one complete set of fields; and a file is a collection of records. A *database management system/software* (see definition) is needed to access information from a database.

database administrator (DBA)

The person who has central control over data and programs accessing the data. Responsibilities of the database administrator include data structure definition, storage and access methods specification, schema and physical organization modification, granting of authorization for data access, and integrity constraint specification. In highway agencies, the database administrator usually belongs to the information systems division, although many functional units have their own DBAs.

database dictionary

A file that defines the basic organization of a database. A database dictionary contains a list of all files in the database, the number of records in each file, and the names and types of each data field. Most database management systems keep the data dictionary hidden from users to prevent them from accidentally destroying its contents. Data dictionaries do not contain any actual data from the database, only bookkeeping information for managing it. Without a data dictionary, however, a database management system cannot access data from the database.

database management system/software (DBMS)

A collection of programs that enables information to be stored in, modified, and extracted from a database. There are many different types of DBMSs, ranging from small systems that run on personal computers to huge systems that run on mainframes. From a technical standpoint, the systems can differ widely. The terms *relational*, *network*, *flat*, and *hierarchical* all refer to the way a DBMS organizes information internally (see *database model*). This internal organization affects how quickly and flexibly the information can be extracted when the DBMS processes requests for information from a database. The requests are made in the form of a query, which is a stylized question (see *database query*).

database model/schema

The structure or format of a database, described in a formal language supported by the database management system. Schemas are generally stored in a data dictionary. Although a schema is defined in text database language, the term is often used to refer to a graphical depiction of the database structure. The following database models are used.

flat file model

A file structure involving data records that have no structured interrelationship. A flat file takes up less computer space than a structured file but requires the database application to know how the data are organized within the file.

Flat File Model

	Route No.	Miles	Activity
Record 1	I-95	12	Overlay
Record 2	I-495	05	Patching
Record 3	SR-301	33	Crack seal

hierarchical model

A data model that links records together like a family tree, but each record type has only one owner (e.g., a purchase order is owned by only one customer). Hierarchical data structures were widely used in the first mainframe database management systems. However, due to their restrictions, they often cannot be used to relate structures that exist in the real world.

Hierarchical Model
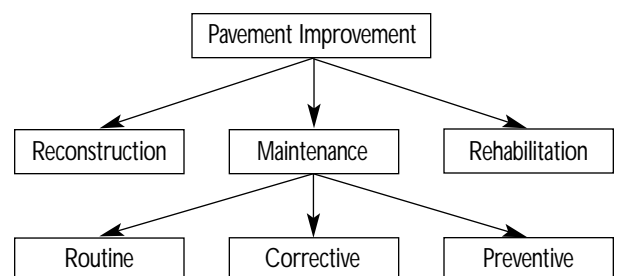
network model

A special case of the hierarchical data model in which each record type can have multiple owners (e.g., purchase orders are owned by both customers and products).

Network Model

Preventive Maintenance
→ Rigid Pavement
→ Flexible Pavement
→ Spall Repair
→ Joint Seal
→ Crack Seal
→ Patching
→ Silicone Sealant
→ Asphalt Sealant

relational model

Data items organized as a set of formally described tables from which data can be accessed or reassembled in many ways without having to reorganize the database tables. Each table (sometimes called a relation) contains one or more data categories in columns. Each row contains a unique instance of data for the categories defined by the columns.

Relational Model

Activity Code	Activity Name
23	Patching
24	Overlay
25	Crack Sealing

Key = 24

Activity Code	Date	Route No.
24	01/12/01	I-95
24	02/08/01	I-66

Date	Activity Code	Route No.
01/12/01	24	I-95
01/15/01	23	I-495
02/08/01	24	I-66

object-oriented model

Defines a data object as containing code (sequences of computer instructions) and data (information that the instructions operate on). Traditionally, code and data have been kept apart. In an object-oriented data model, the code and data are merged into a single indivisible thing—an object.
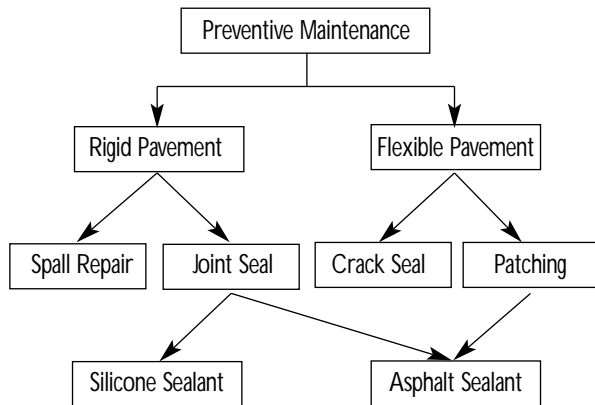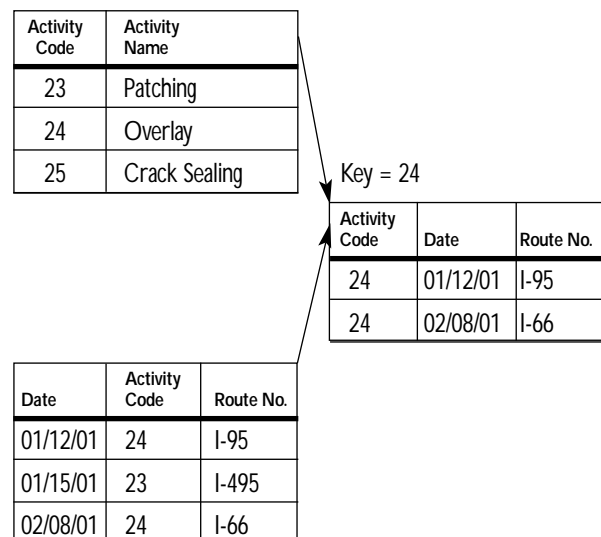
Object-Oriented Model

Object 1: Maintenance Report

Date	
Activity Code	
Route No.	
Daily Production	
Equipment Hours	
Labor Hours	

Object 1 Instance

01-12-01
24
I-95
2.5
6.0
6.0

Object 2: Maintenance Activity

Activity Code	
Activity Name	
Production Unit	
Average Daily Production Rate	

database query

A request for information from a database. There are three general methods for posing queries: (1) Choosing parameters from a menu: In this method, the database system presents a list of parameters from which to choose. This is perhaps the easiest way to pose a query because the menus provide guidance, but it is also the least flexible; (2) Query by example (QBE): In this method, the system presents a blank record and lets the user specify the fields and values that define the query; and (3) Query language: Many database systems require the user to make requests for information in the form of a stylized query that must be written in a special query language. This is the most complex method because it forces the user to learn a specialized language, but it is also the most powerful.

database replication

The process of duplicating a portion of a database from one environment to another and keeping the data in other environments consistent with the data in the source database.

database server

Often used interchangeably to refer to hardware or software. Both uses pertain to the same principle—a database architecture prepared to receive requests from a third party, or client, and respond to those requests by delivering a particular type of information. In either case, appropriate software is the core of the system. Referring to a piece of hardware as a "server" usually implies that it is running one or more pieces of server software.

decision support system (DSS)

A generic term used to refer to a computer system or software that supports data analysis, reporting, and other data processing capabilities to help an organization or enterprise (see definition of *enterprise*) make informed decisions.

desktop application

An information processing and analysis tool that accesses or queries the source database or data warehouse across a network using an appropriate database interface. The desktop application manages the human interface between data sources and data users.

distributed database

A database that consists of two or more data files located at different sites on a computer network. Because the database is distributed, different users can access it without interfering with one another (see also *interoperable database*). In transportation agencies, distributed databases enable multiple individuals in separate locations to run different applications using the same data. However, the database administrator or manager needs to periodically synchronize the scattered databases to make sure they have consistent data.

document management

Keeping track of stored documents that are either (1) created by a computer using word processing, spreadsheet, or other software; (2) scanned into a computer and converted to computer code by optical character recognition software; or (3) scanned into a computer and saved in a computer image format. Document management systems provide the ability to store different documents in one location and view them from any location on the computer network, with all the benefits of controlled backup, recovery, and disaster management protection.

dynamic segmentation

A method of modeling linear features for various applications including highway analysis. Dynamic segmentation is a process for determining the locations of events on linear features (e.g., highway segments) at run time, straight from the tables of features for which distance measures are available and without changing the underlying data structure. For transportation infrastructure management, dynamic segmentation provides the ability to integrate asset data that commonly use linear location measurements (e.g., mileposts).

enterprise

An organizational unit or entity made up of business functions, divisions, or other components that perform various responsibilities to achieve a common objective. The terms "enterprise database" and "enterprise modeling" are used, respectively, to describe data and the method used to understand their relationships across an enterprise environment.

enterprise application integration (EAI)

Refers to extraction, transformation, and loading of data from *enterprise resource planning (ERP)* (see below) systems to other data sources. Software tools that perform this function are generally batch or point-in-time products designed for initial or ongoing loading of data warehouses.

enterprise resource planning (ERP)

An integrated, technology-driven approach to managing enterprise resources, whether they are cash, raw materials, or personnel. ERP is a strategic tool to help an organization integrate all its business processes and data.

entity

In programming, engineering, and many other contexts, "entity" is used to refer to units, whether concrete things or abstract ideas, that have no specific names. One can draw something and refer to that drawing as the representation of an entity.

entity relationship (ER) diagram

A design tool that graphically expresses the database's overall logical structure by representing data entities, their cardinality relationships (e.g., one-to-one, one-to-many, etc.), and the linkages that connect them. The construction of an ER diagram is essential for the design of database tables, extracts, and metadata (see definition of *metadata*). ER diagrams are especially important in designing and developing data documentation and tables that represent data relationships (see *data/process flow diagram* and *data structure diagram*).

Entity Relationship Diagram

(shows relationship of data elements)
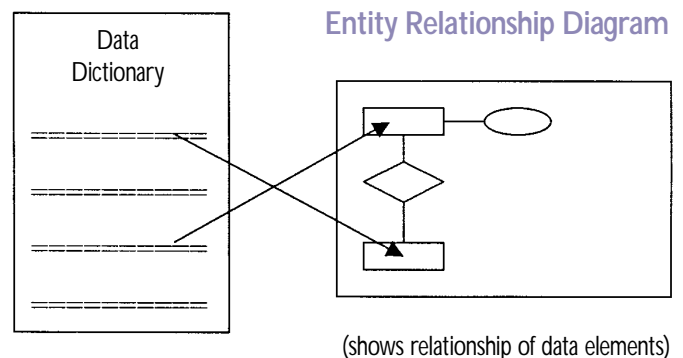
executive information system (EIS)

A computerized system or tool programmed to provide predefined reports or summary information to chief administrators or high-level executives. An EIS provides powerful information reporting and drill-down capabilities, including ad hoc querying and other analytical processing functions.

extensibility

The ability to incorporate new functionality in existing database systems without making major software changes or redefining the basic system architecture.

geographic information system (GIS)

A computer-based tool used to gather, transform, manipulate, analyze, and produce information related to the surface of the earth. This data may exist as maps, three-dimensional models, tables, or lists. A GIS can be as complex as a whole system that uses dedicated databases and workstations hooked up to a network, or as simple as off-the-shelf desktop software. Transportation agencies use GIS for a variety of infrastructure management purposes including mapping and location identification, spatial analysis, presentation, and reporting of inventory and other useful information. The GIS is an important data integration tool for Asset Management because every asset can be tied or referenced to a particular location on a highway map and analyzed using GIS software (see *location reference system*).

graphical user interface (GUI)

A computer program interface that takes advantage of a computer's graphics capabilities to make a program easier to use. Well-designed GUIs can free the user from learning complex command languages, as well as make it easier to move data from one application to another. A true GUI includes standard formats for representing text and graphics. Because the formats are well-defined, different programs that run under a common GUI can share data. This makes it possible, for example, to copy a graph created by a spreadsheet program into a document created by a word processor. Typical Windows applications have GUIs. Many Disk Operating System (DOS) programs include some features of a GUI, such as menus, but are not graphics based. Such interfaces are sometimes called "graphical character-based user interfaces" to distinguish them from true GUIs.

information systems architecture

An information systems framework that describes: (1) business rules—functions a business performs and the information it uses; (2) system structure—definitions and interrelationships between applications and the products, (3) technical specifications—the interfaces, parameters, and protocols used by the product and applications; and (4) product specifications—standards pertaining to the elements of the technical specifications and the application of vendor-specific tools and services in developing and running applications.

interoperable database

Also called "federated database." A collection of autonomous and possibly heterogeneous database systems over multiple sites that are connected through a communications network (see also *distributed database*).

Java

An object-oriented computer programming language (see also *object-oriented programming*) based on C++ but without many of its rarely used features. Java was designed to support applications on networks with a variety of computer processing and operating system architectures, making the compiled Java applications executable anywhere on many processors on the network. Java is ideal for a diverse computing environment like the Internet.

legacy system

An older computer system or application program that an agency continues to use. Many legacy systems are unable to meet the changing business processes of organizations, thereby presenting a significant challenge for data integration.

local access database (LAD)

Also called "data mart." A database that serves individual systems and workgroups as the end point for shared data distribution. LADs are the "retail outlets" of data warehouse networks. They provide direct access to the data requested by specific systems or desktop query services. Data are propagated to LADs from data warehouses according to orders for subsets of certain shared data tables and particular attributes therein, or subsets of standard collections. The propagated data are usually located on a LAN server.

location reference system

A system for storing, maintaining, and retrieving location information. One technique is to locate a specific position with respect to a known point. While spatial features are typically located using planar (two-dimensional) referencing systems like geographic coordinates, many transportation features—roads, bridges, and other structures—are located using linear (one-dimensional) referencing systems including the route-milepost system. A linear reference system cannot entirely replace a planar referencing system for geographic

display. However, it does represent the format in which most transportation infrastructure data are currently stored and reported.

location transparency

A mechanism that keeps the specific physical address of data or an object unknown to the user. This is done by resolving the location of the data within the system so that operations on the data can be performed without knowledge of its actual physical location.

metadata

Information that describes or characterizes data. Metadata are used to provide documentation for data products. In essence, metadata answer the who, what, when, where, why, and how about every facet of the data that are being collected and documented.

mini mart

A small subset of a data warehouse used by a small number of users. The mini mart is a very focused slice of a larger data warehouse.

MIPS

Acronym for "millions of instructions per second." MIPS is sometimes mistakenly considered a relative measure of processing capability among computer models and products, but it is a meaningful measure only among versions of the same computer processors configured with identical peripherals and software.

multidimensional database (MDB)

A type of database that is optimized for data warehousing and online analytical processing (OLAP) applications (see *online analytical processing*). Multidimensional databases are frequently created using input from existing relational databases. MDB denotes the ability to process the data in the database quickly so that answers can be generated immediately.

object-oriented programming (OOP)

A method in which programmers define not only the type of data structure, but also the types of operations (functions or procedures) that can be applied to the data structure. In this way, the data structure becomes an object that includes both data and functions. In addition, programmers can create relationships between one object and another. For example, objects can inherit characteristics from other objects. One of the principal advantages of OOP techniques over traditional procedural programming techniques is that they enable programmers to create modules that do not need to be changed when a new type of object is added. A programmer can simply create a new object that inherits many of its features from existing objects, making object-oriented

programs easier to modify. To perform object-oriented programming, one needs an object-oriented programming language. Java, C++, and Smalltalk are three of the more popular OOP languages, and there are also object-oriented versions of Pascal (see also *database model/schema: object-oriented model*).

online analytical processing (OLAP)

A category of software tools that provides analysis of data stored in a database. OLAP tools enable users to analyze different dimensions of multidimensional data. For example, it provides time-series and trend-analysis views. The chief component of OLAP is the OLAP server, which sits between a client and a database management system. The OLAP server understands how data is organized in the database.

online transaction processing (OLTP)

A type of computer processing in which the computer responds immediately to user requests. Each request is considered to be a transaction. Automatic teller machines are examples of transaction processing. The opposite of transaction processing is batch processing, in which a batch of requests is stored and then executed all at one time. Transaction processing requires interaction with a user, whereas batch processing can take place without a user being present.

Open Database Connectivity (ODBC)

A standard database access method developed to make it possible to access any data from any application, regardless of which *database management system (DBMS)* (see definition) is handling the data. ODBC does this by inserting a middle layer, called a database driver, between a database application and the DBMS. The purpose of the driver is to translate the application's data queries into commands that the DBMS understands. For this to work, both the application and the DBMS must be ODBC-compliant; that is, the application must be capable of issuing ODBC commands and the DBMS must be capable of responding to them. ODBC is an essential data integration element.

open systems environment

A computing environment that allows users to access and utilize data and software across multiple platforms. Incompatible hardware platforms, operating systems, and other application software are tied together through the use of industry-standard system components. An open systems environment facilitates systems integration by creating the means for resource sharing in an environment in which hardware, software, and data can be efficiently accessed and applied by all users in an organization.

Open Systems Interconnection (OSI) Standard

The computer industry standard network protocol or model for computer-to-computer communications. Developed by the International Standards Organization, OSI consists of seven layers or modules arranged in a hierarchy (see below) with each layer performing a function that is dependent on the more elementary (lower) layer. The system is "open" in the sense that layers are not tightly specified, which allows different networks to easily link or interconnect.

OSI Layer	Function
1. Physical	Activation, maintenance, and deactivation of physical connection.
2. Data Link	Synchronization, error detection and recovery, and flow control for information.
3. Network	Establishment, maintenance, and termination of switched connections.
4. Transport	Selection of network service. Data flow regulation between end users.
5. Session	Provision to transfer data and control in an organized and synchronized manner.
6. Presentation	Delivery of information to the end users in usable and understandable forms.
7. Application	Facility to serve end user. Authentication of user and destination identifications. Authority to exchange information.

prototype application/prototyping

A software development process whose purpose is to evaluate the suitability of a potential application for production.

rapid application development (RAD)

A method of building computer systems in which the system is programmed and implemented in segments, rather than waiting until the entire project is completed for implementation. RAD uses such tools as CASE (see *Computer Aided Software Engineering Tools*) and *visual programming* (see definition).

real time

A degree of computer responsiveness that a user considers to be adequately immediate or that allows the computer to keep pace with some external process.

record

A collection of data items arranged for processing by a program. Multiple records are contained in a file or data set. The organization of data in the record is usually prescribed by the programming language that defines the record's organization or by the application that processes it. Typically, records can be of fixed or variable length with the length information contained within the record.

redundancy

The storage of multiple copies of identical data. The process of limiting excessive copying, update, and transmission costs associated with redundant data is called "redundancy control." *Database replication* (see definition) is a strategy for redundancy control with the intention to improve system performance. "Redundancy" may also be used to refer to backup systems that take over data processing and transmitting functions when a primary system fails.

relational database management system

A database or database management system that stores information in tables—rows and columns of data—and conducts searches by using data in specified columns of one table to find additional data in another table. In a relational database, the rows of a table represent records (collections of information about separate items) and the columns represent fields (particular attributes of a record). In conducting searches, a relational database matches information from a field in one table with information in a corresponding field of another table to produce a third table that combines requested data, (see also *database model/schema: relational model*).

repository

A central place in which databases are stored and maintained in an organized way. A repository may be directly accessible to users or it may be a place from which specific databases, files, or documents are obtained for further relocation or distribution in a network.

reverse data engineering

Also called "data reengineering." Allows the user to capture physical models of legacy and production systems and relate each attribute in the data model to the database(s), table(s), and column(s) from which it is derived. Data reengineering is suitable for organizations that need to augment an existing system, particularly if the system requires new database designs, new screen designs, or new programs. It is particularly useful for designing a data warehouse and an excellent tool for a database manager who has inherited a system that is partially or totally undocumented.

scalability

The ability to change size to support larger or smaller volumes of data and more or fewer users with minimal impact on the unit cost of business and the procurement of additional services.

spatial data

Any information about the location, shape, and relationships of geographic features. For highways this includes the relative locations and distances of transportation infrastructure.

standards

Statements of fact, quality, procedures, or content, to which applicable data entities are compared for purposes of acceptance or use. Standards are documented agreements used to justify decisions, implement policy, and ensure that data processes, products, or services meet their intended purpose. Using common standards increases the shareability, reliability, and effectiveness of data. Many different standards exist that pertain to databases, communications, and computer network procedures, all very essential to data integration. Standards are prescribed by several standard-setting organizations such as American National Standards Institute, International Standards Organization, and National Institute for Standards and Technology.

stored procedure

A set of structured query language (SQL) statements (see definition of *structured query language*) that is stored in the database with an assigned name and in compiled form so it can be shared by a number of programs. The use of stored procedures can be helpful in controlling access to data (end users may enter or change data but not write procedures), preserving data integrity (information is entered in a consistent manner), and improving productivity (statements in a stored procedure need to be written only once).

structured query language (SQL)

A standardized language for querying information from a database. SQL was first introduced as a commercial database system in 1979 and has since been the favorite query language for database management systems running on minicomputers and mainframes. Increasingly, however, SQL is being supported by PC database systems because it supports distributed databases (see definition of *distributed database*). This enables several users on a computer network to access the same database simultaneously. Although there are different dialects of SQL, it is nevertheless the closest thing to a standard query language that currently exists.

system administrator

An individual responsible for maintaining a multiuser computer system such as a computer network. The duties of the system administrator typically include adding and configuring new workstations, establishing user accounts, installing system-wide software, performing computer virus protection procedures, and allocating computer storage space. The system administrator is often called the "sysadmin." Small organizations usually have only one system administrator, whereas larger enterprises with complex computer network architecture may have a whole team of system administrators.

target database

The database in which data will be loaded or inserted.

Unified Modeling Language (UML)

A standard analysis and design language created to assist in modeling complex software systems by using a common object-oriented notation. UML is rapidly becoming the de facto standard for describing and sharing system design data. Many software vendors have chosen UML as their notation for repository (see definition of *repository*) products to aid in sharing common software components among different third-party modeling and programming tools.

use case analysis

A methodology used in systems analysis to identify, explain, and organize system requirements. It is made up of a set of likely sequences of interactions between systems and users in a specific environment and associated with a particular goal. Use cases can be applied at various stages of software development including planning, software design, development, and testing.

versioning

The maintenance and storage of previous copies of a piece of information for security, diagnostics, and other interests. Versioning also pertains to the ability to create and manage different versions of the same information.

very large database (VLDB)

Sometimes used to describe databases occupying magnetic storage in the terabyte (1 trillion bytes) range and containing billions of table rows. Typically, these are decision-support systems or transaction processing applications serving a large number of users.

visual programming language

A programming language that uses, partially or entirely, visual representations such as graphics, icons, drawings, and animation. A visual language handles visible information, supports visual interaction, and allows programming of visual expressions.

workflow management

Based on the concept that business processes are actually sets of tasks done in a prescribed order (workflow) that combines information from various sources. Workflow management focuses on managing the flow of information as it is processed, shared, manipulated, and compiled from one participant to another in a way that is governed by rules or procedures. Integration of asset management data is an example of a workflow management procedure in State transportation agencies.

For further information on data integration initiatives of the FHWA Office of Asset Management, contact:

Office of Asset Management
Federal Highway Administration
U.S. Department of Transportation
400 Seventh Street, S.W., HIAM-30
Washington, DC 20590

Telephone: 202-366-9242
Fax: 202-366-9981

U.S. Department
of Transportation

FHWA IF-01-017