Unit #8

Database Design MIS5203

Agenda

- Case Study Database design with input from legacy information systems
- Database Design Process / Data Normalization
- Reading a Database Design
- Database Design Security Considerations
- Database Design Assignment Review
- Quiz

Case Study – MUDRA Communications

What legacy database management systems (DBMS) were MUDRA using?

Period	Hardware	OS	Application	Basis	Reasons for change
1982-86	Zenith CPM	DOS	Financial Accounting	Cobol	Third-party software; only
	based		Systems (FAS)		one location
		DOS	FAS – Dbase	In-house	DOS was popular
1986-87	Pentium	Windows	POP – Clipper & Dbase	In-house	Entry-level systems
1988-90	Horizon III	Unix	FAS – Unify database	In-house	Increase in volume
	Pentium	Windows	Clipper and Dbase		
1992-94	Wipro-	Unix	FAS – Ingres	In-house	Vendor support was
	Landmark				inadequate
	Sun Solaris	Sun	FAS – Ingres database		Increase in volume
	Sun Solaris	Sun	CSA – Ingres database	In-house	
1994-2001	HP Servers	Unix	For Mumbai, Delhi and	Distributed	FAS on Ingres
	(3)		Ahmedabad offices	computing	

What "new" database management systems (DBMS) were MUDRA considering?

Proposed					
	IBM Servers	Linux	FAS – Oracle 8i and 9i	Central	In-house development
2001-2011	(6)			server	
	Dell Servers	Linux	CAS	Central	In-house development
				server	
	Dell Servers	Linux	Varied		In-house development

Case Study – MUDRA Communications

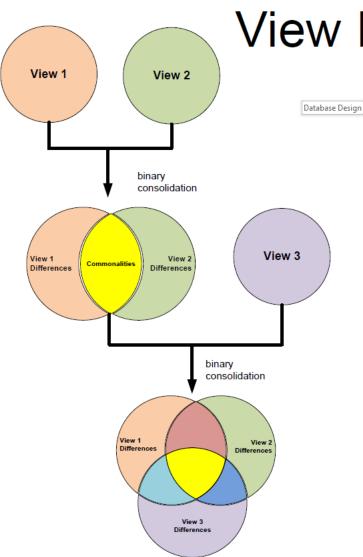
What challenges was MUDRA facing in migrating their legacy databases to a new database management platform?

- Indian law required all commercial legacy data be retained for at least 8 years
- Joseph needed to decide if all historical data would be migrated to the ES or preserved in its current form and extracted manually when required by law
- Coming up with an approach for doing the data migration

What is the essence of the technical problem of data migration?

Database Design

How would you approach migrating legacy databases into new database design model?

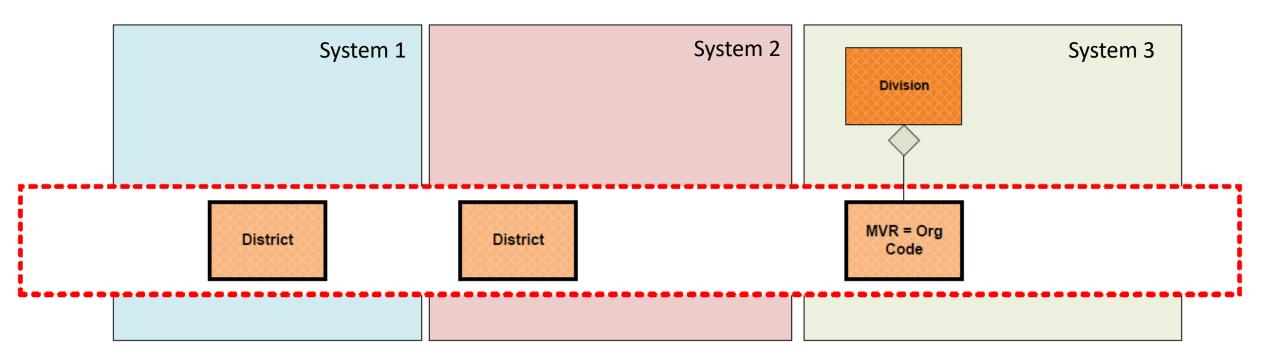


View Integration

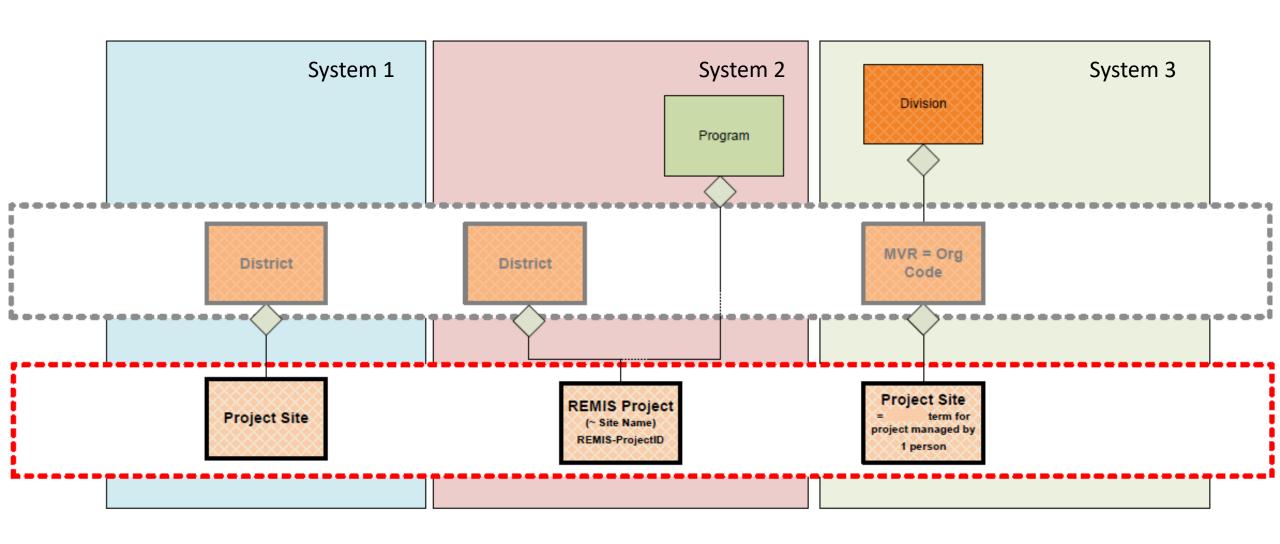
Data model integration achieved by:

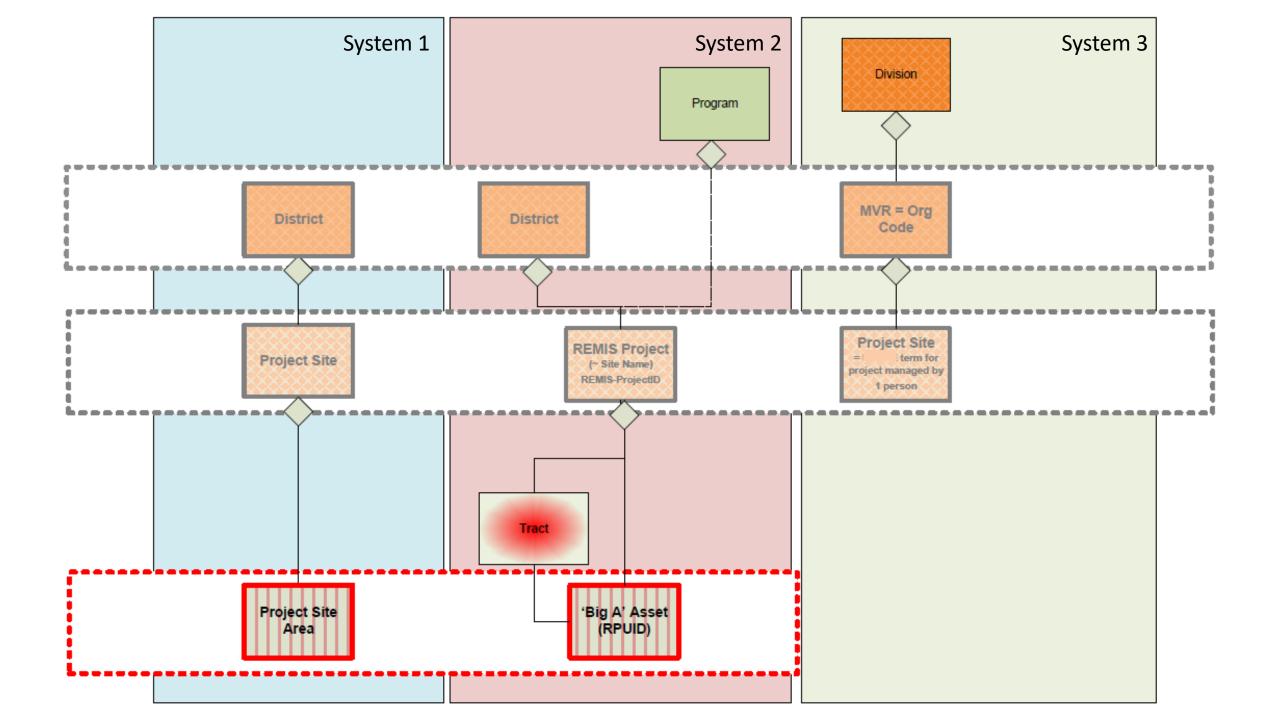
- 1. Identifying,
- 2. Resolving, and
- 3. Consolidating
 - Commonalities (and synonyms)
 and
 - **Differences** (and homonyms)

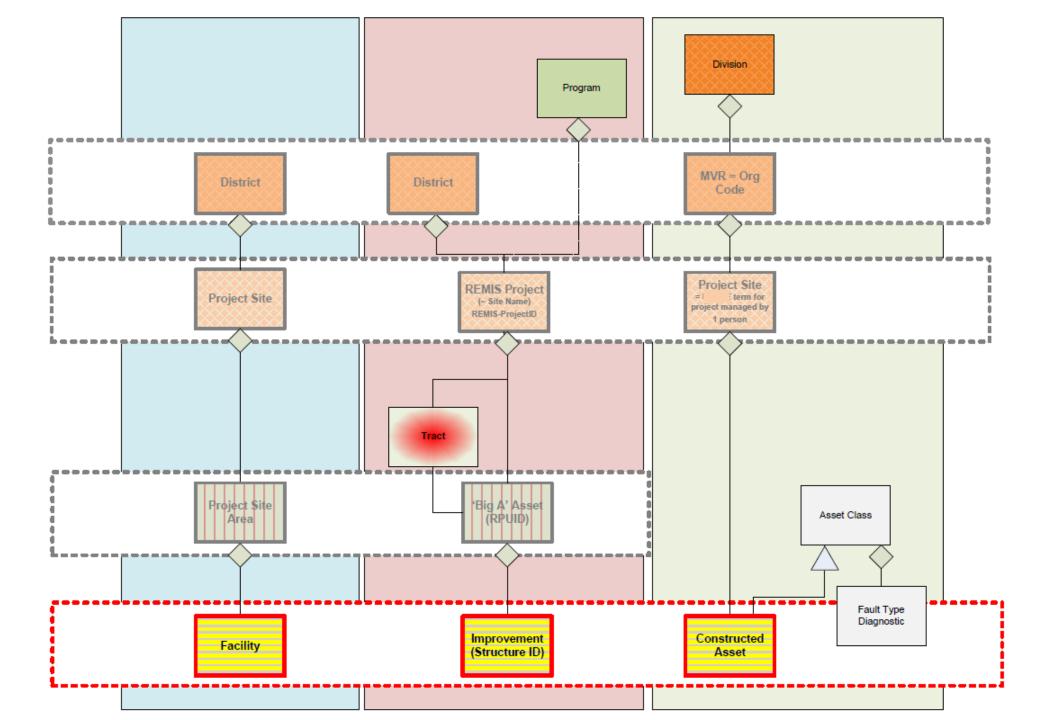
Analysis of data commonalities among legacy information systems

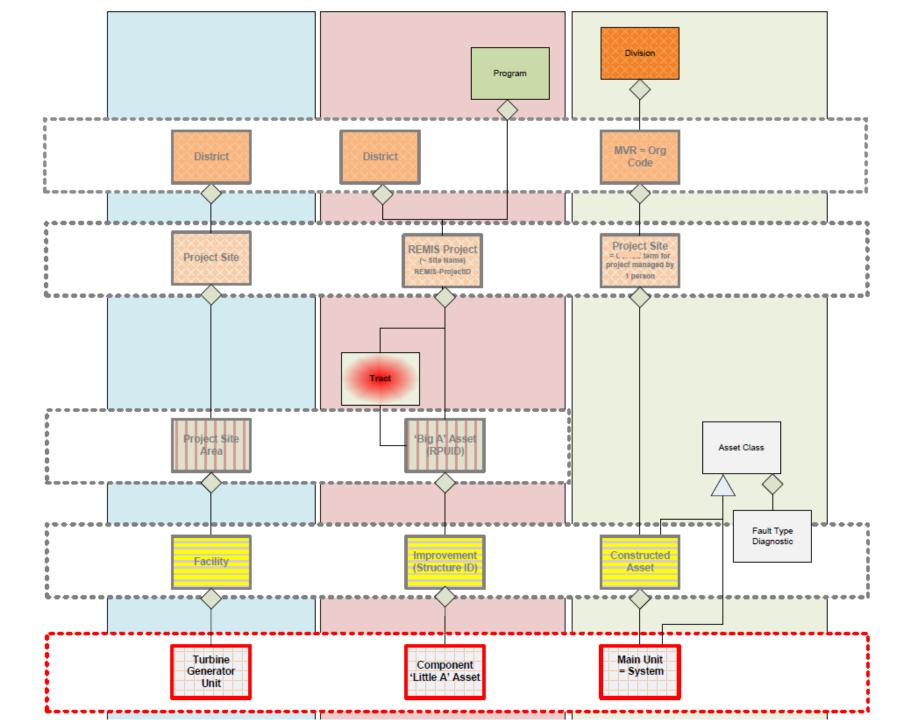


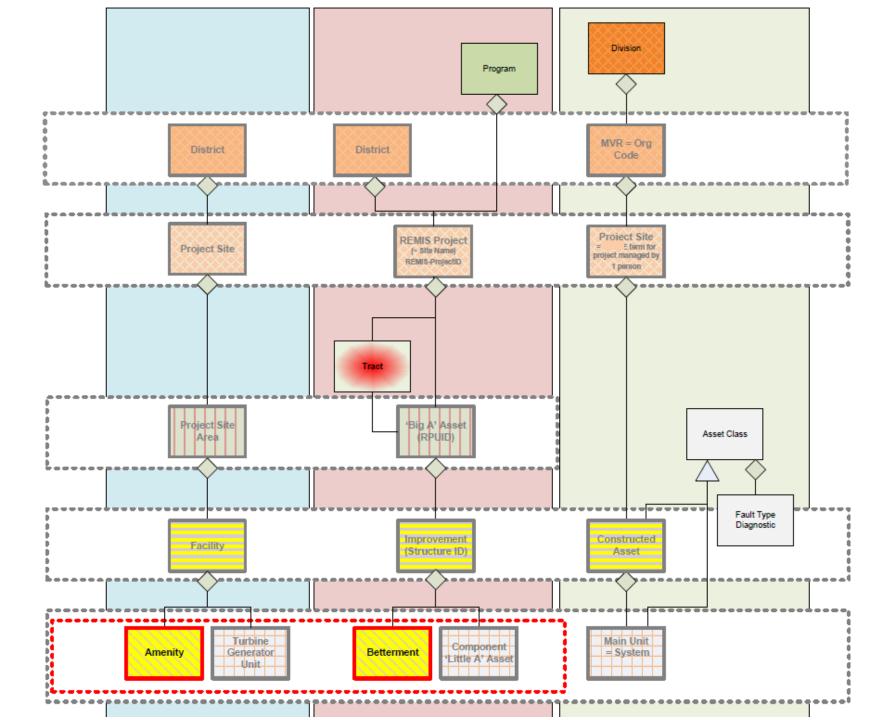
Analysis of data commonalities among legacy information systems

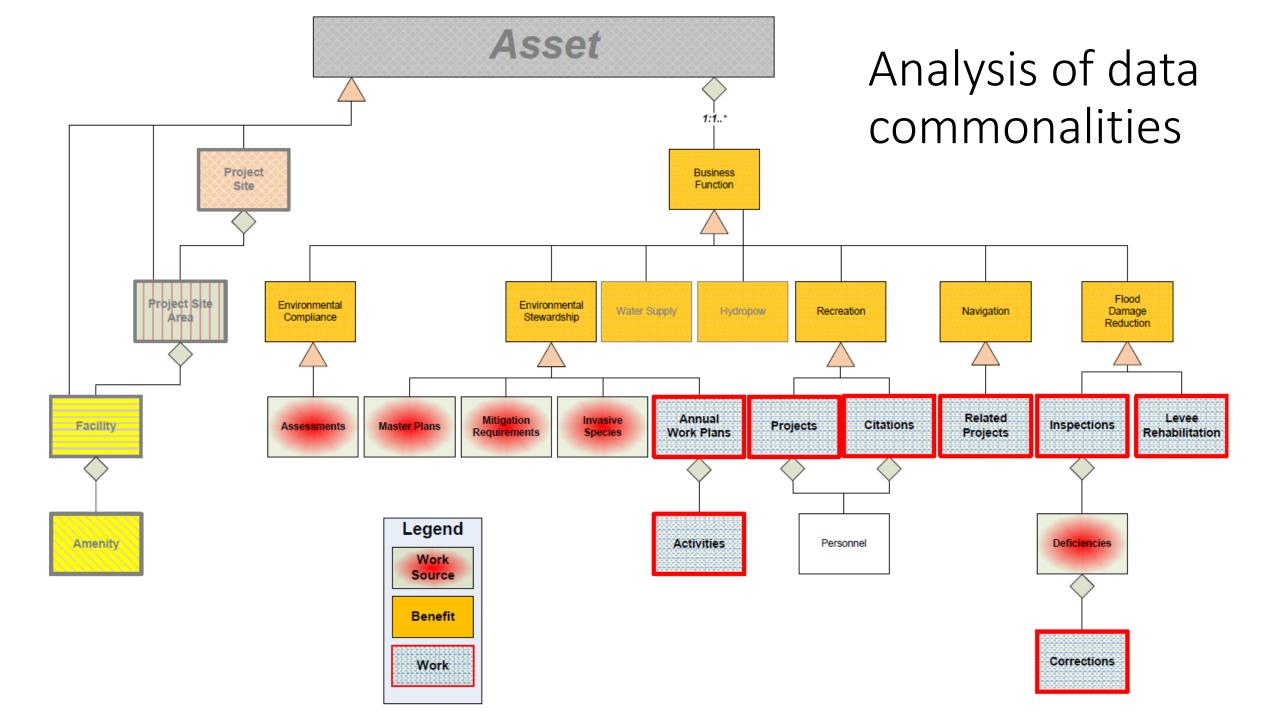






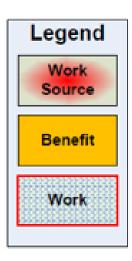






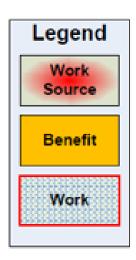
Program District **REMIS Project** (~ Site Name) REMIS-ProjectiD Work Item (P2) Task (P2) Costs Improvement (Structure ID) Component FTE Material Costs Costs

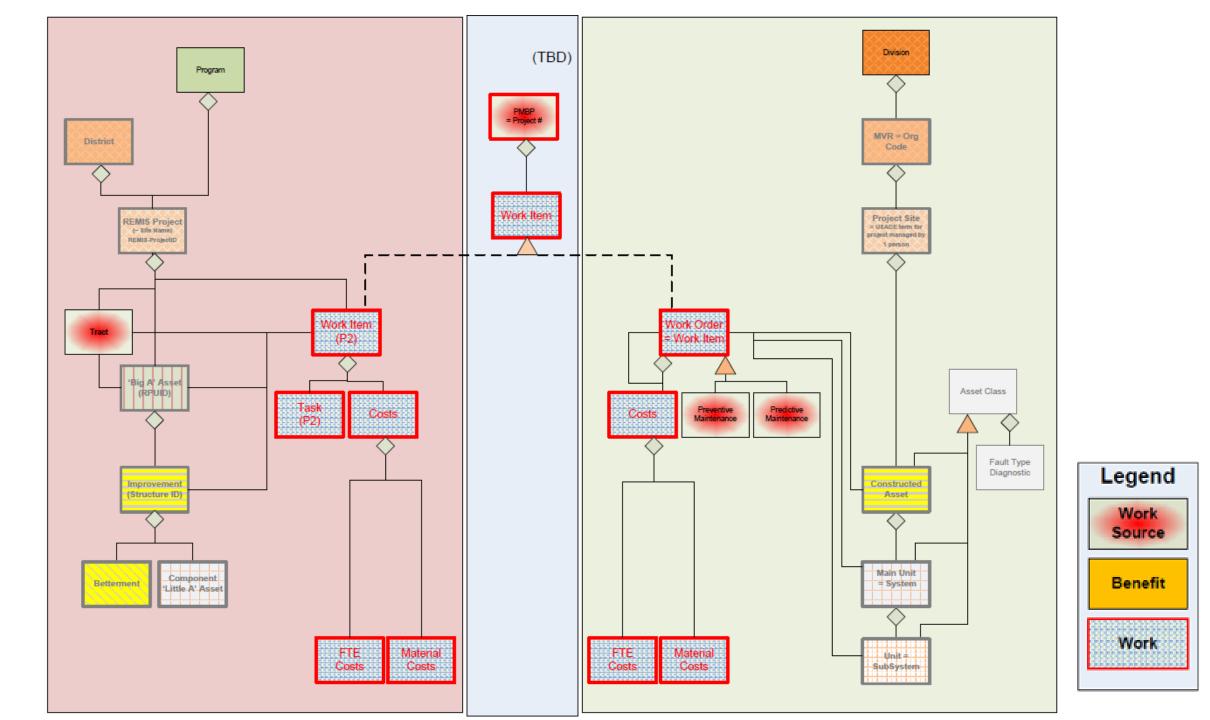
Analysis of data commonalities

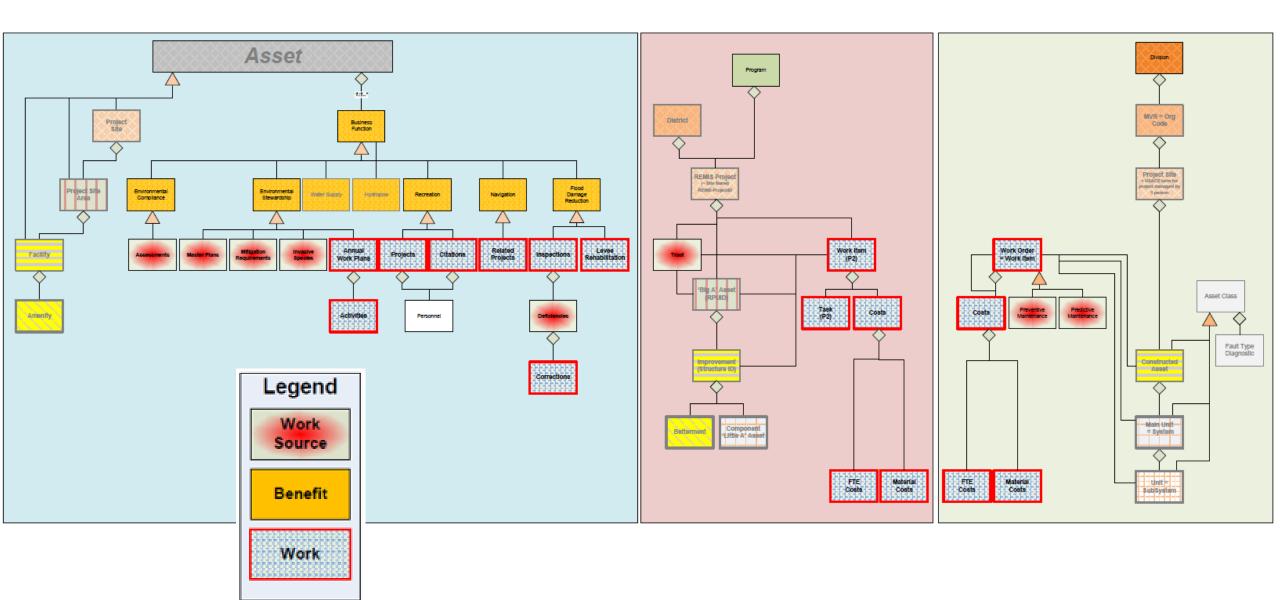


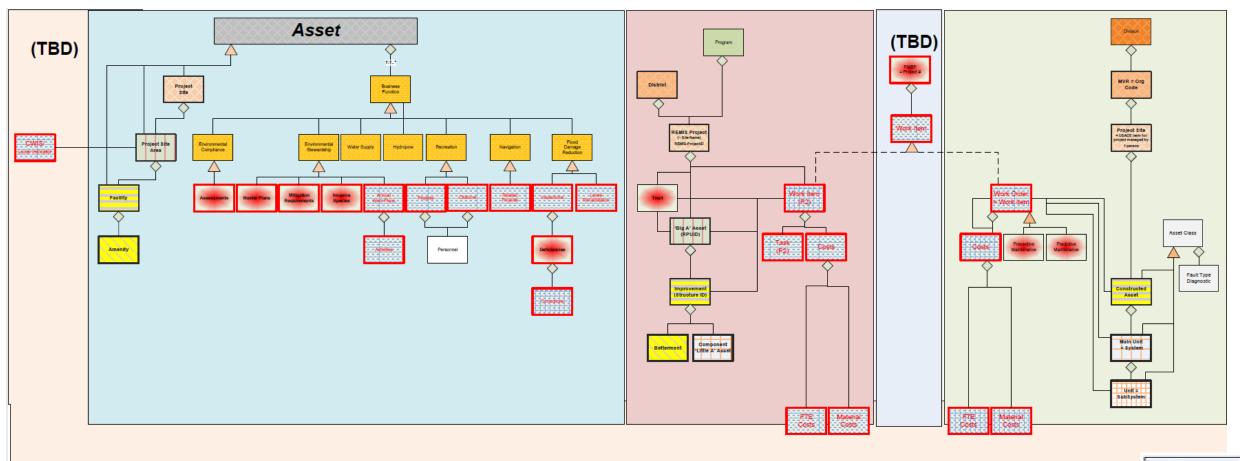
Division MVR = Org Code Project Site = USACE term for project managed by 1 person Work Order = Work Item Asset Class Costs Maintenance Fault Type Diagnostic Constructed Main Unit = System Material FTE Costs Costs

Analysis of data commonalities



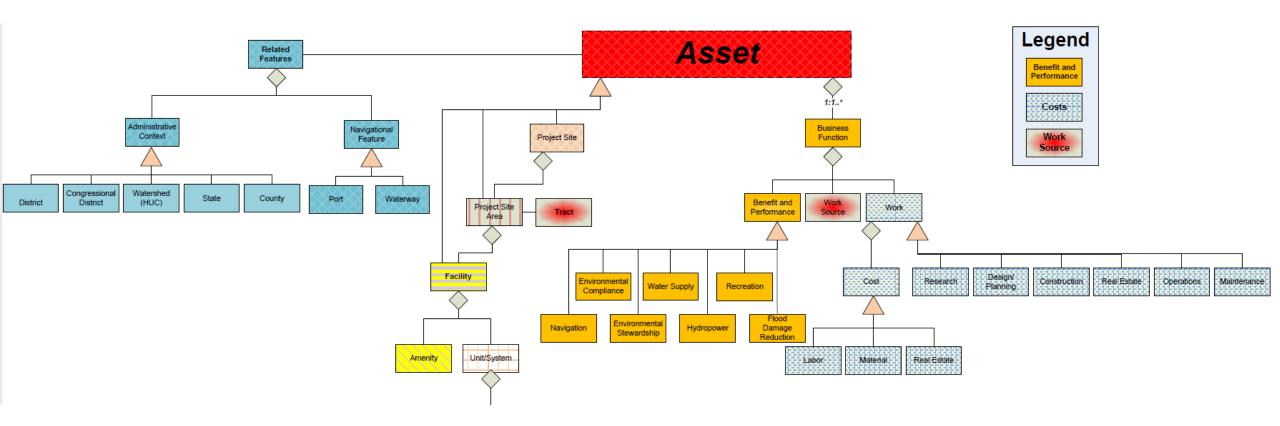




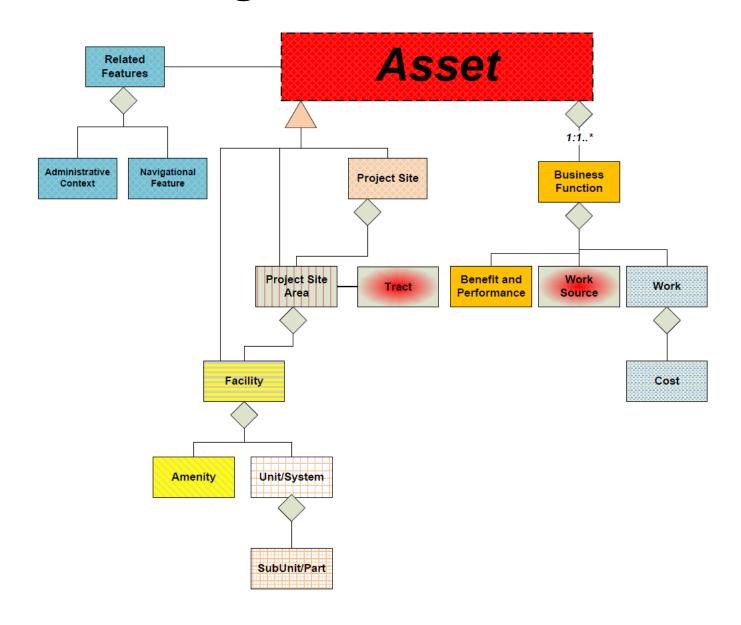




New database design for consolidation



New database design for consolidation



Agenda

- ✓ Case Study Database design with input from legacy information systems
- Database Design Process / Data Normalization
- Reading a Database Design
- Database Design Security Considerations
- Database Design Assignment Review
- Quiz

Database Design Process

The SDLC cycle includes database analysis activities in all phases

- Entity-Relationship (E-R) and Data Flow modeling dominate planning and analysis phases
- E-R models are inputs to the logical database design, often
 - Translated into attribute tables
 - Linked via primary and foreign key relationships
 - Structured within relational database management systems (RDBMS)

 Conceptual data mode (E-R with only entities for specific project) Planning · Conceptual data Data model Maintenance **Analysis** models (E-R with evolution attributes) Implementation Design Logical data model (relational) Database and file definitions (DBMS specific code) and physical file and database design (file organizations)

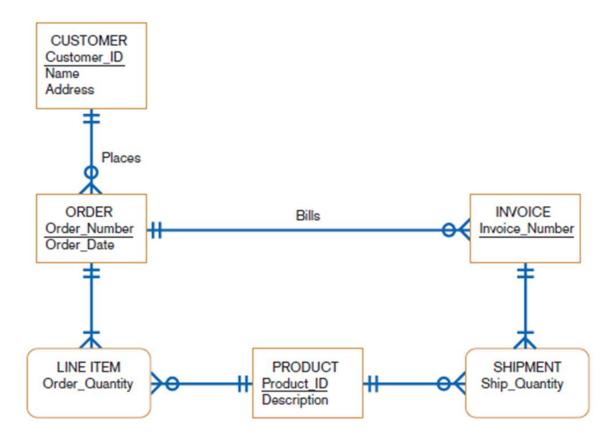
Enterprise-wide data model (E-R with only entities)

Database Design Process - deliverables

- Logical database design
 - Accounts for every data element input and output by an information system
 - Normalized tables and relationships are the primary deliverable
- Physical database design
 - Tables and Views
 - Relationships
 - Attribute domains
 - Specified in
 - Relational database model
 - Accompanying Data Dictionary

"Relation" = a named, two-dimensional table of data

 Each relation consists of a set of named columns and an arbitrary number of unnamed rows



Delationer

CUSTOMER(Customer_ID, Name, Address)
PRODUCT(Product_ID, Description)
ORDER(Order_Number, Customer_ID, Order_Date)
LINE ITEM(Order_Number, Product_ID, Order_Quantity)
INVOICE(Invoice_Number, Order_Number)
SHIPMENT(Invoice_Number, Product_ID, Ship_Quantity)

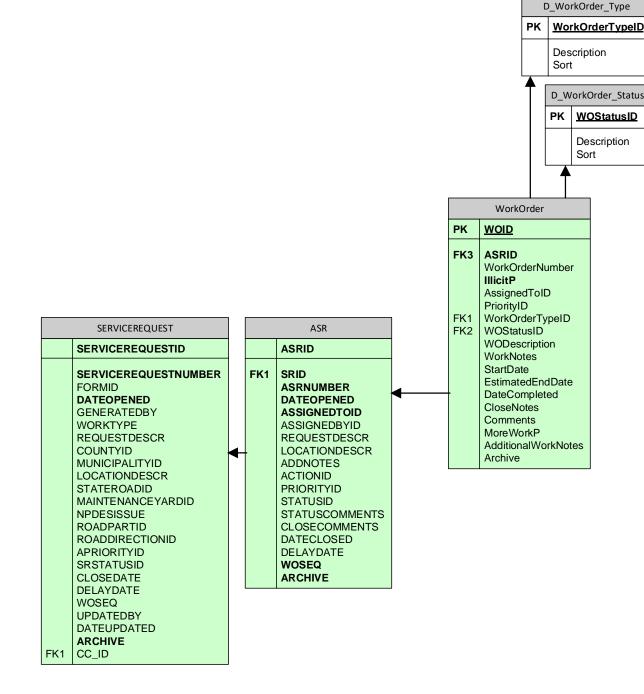
Relational Tables

Table

- A relation that contains a minimum amount of redundancy
- Allows users to insert, modify, and delete the rows without errors or inconsistencies

Primary Key

- An attribute whose value is unique across all occurrences of a relation
- All relations have a primary key
 - Ensures the rows (i.e. data records) are unique
 - A primary key often in a single attribute, but may be composed of multiple attributes ("compound key")



Data Normalization

A structured process for organizing data into database tables in such a way that it preserves the relationships among the data

First Normal Form (1NF)

- Unique rows, no multivalued attributes
- All tables have primary keys. B
- "No multivalued attributes" means a table cannot have a field with multiple values. But, as
 we saw in chapter 8, it's possible to define an entity with a multivalued attribute (such as
 an employee having many skills). Therefore, the original entity may need to be
 reconstructed into more than one relation.
- All relations are in 1NF

Second Normal Form (2NF)

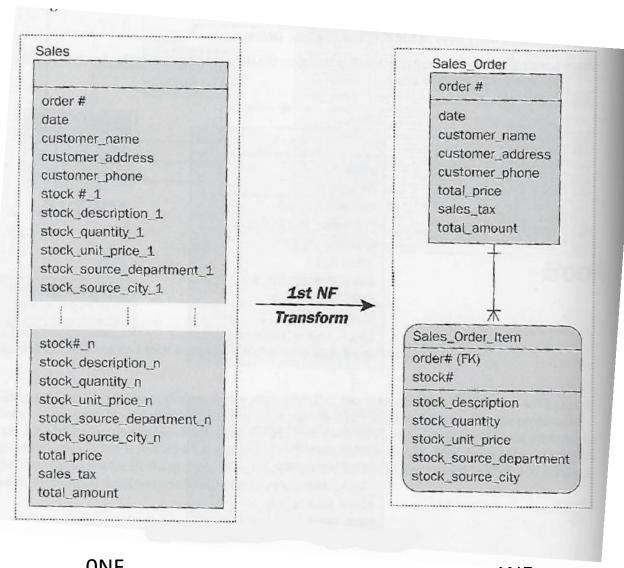
Each nonprimary key attribute is identified by the whole primary key (called full functional dependency)

Third Normal Form (3NF)

Nonprimary key attributes do not depend on each other (i.e. no transitive dependencies)

First Normal Form (1NF)

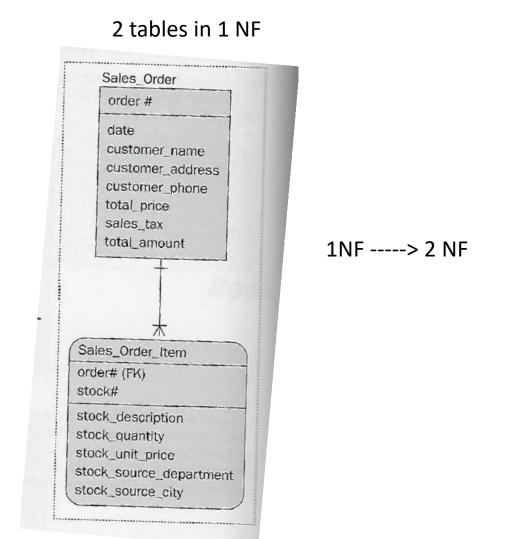
- Unique rows, no multivalued attributes
 - All tables have primary keys
 - "No multivalued attributes" means
 - Table cannot have a field with multiple values
 - Need to reconstruct this information with more than one table
- All relations (i.e. tables) are in 1NF



ONF 1NF

Second Normal Form (2NF)

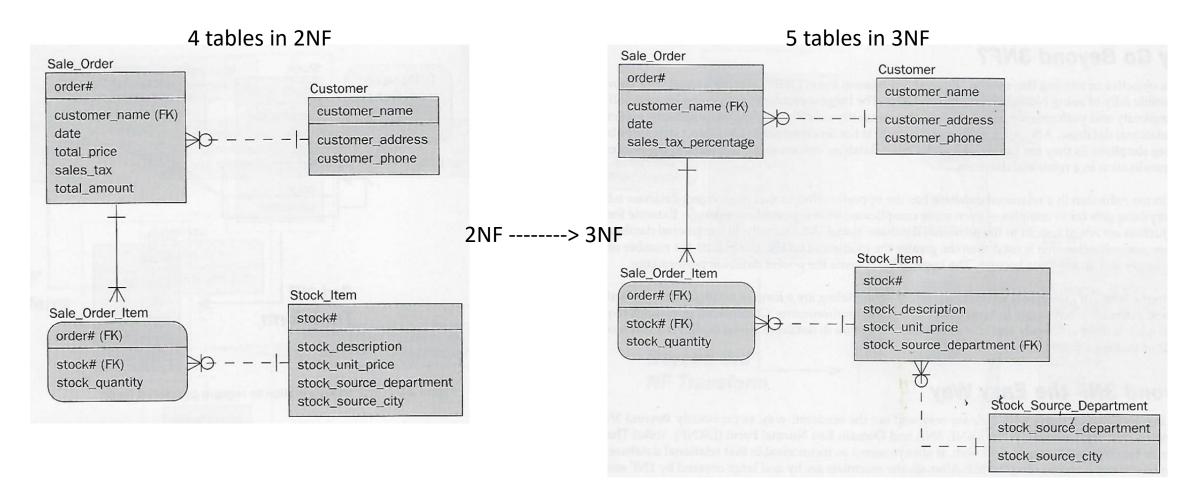
• Each nonprimary key attribute is identified by the whole primary key, called full functional dependency



4 tables in 2 NF Sale_Order order# Customer customer name customer name (FK) date customer address total price customer_phone sales_tax total amount Stock_Item Sale Order Item stock# order# (FK) stock_description stock# (FK) stock_unit_price stock_quantity stock_source_department stock_source_city

Third Normal Form (3NF)

Nonprimary key attributes do not depend on each other (i.e. no transitive dependencies)

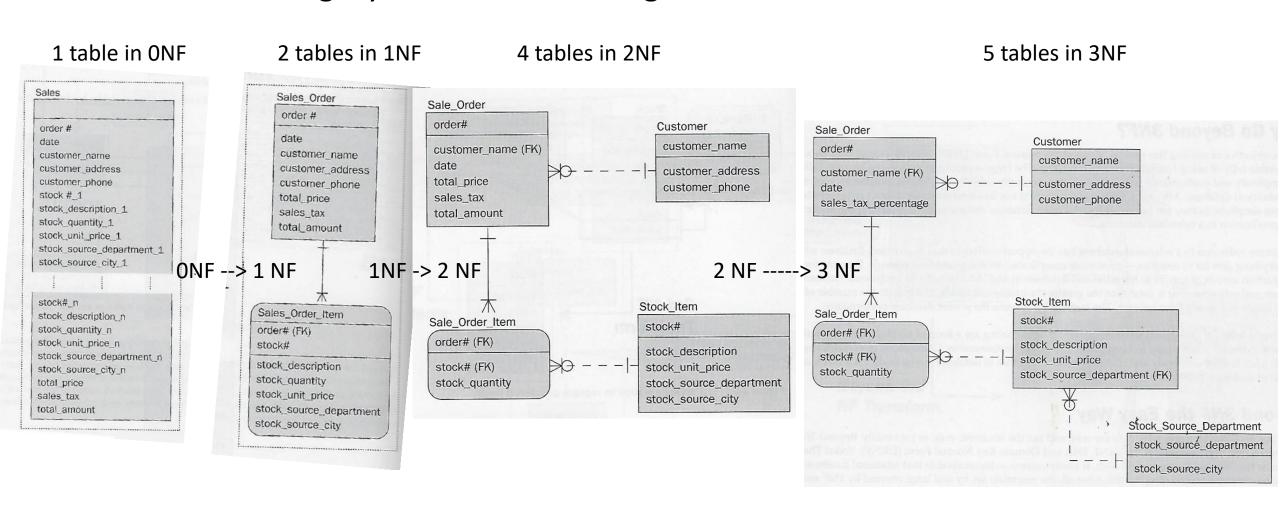


Third Normal Form (3NF) continued

- Foreign Key: an attribute that appears as a nonprimary key attribute (or part of a primary key) in one relation and as a primary key attribute in another relation
- **Referential Integrity:** an integrity constraint specifying that the value (or existence) of an attribute in one relation depends on the value (or existence) of the same attribute in another relation

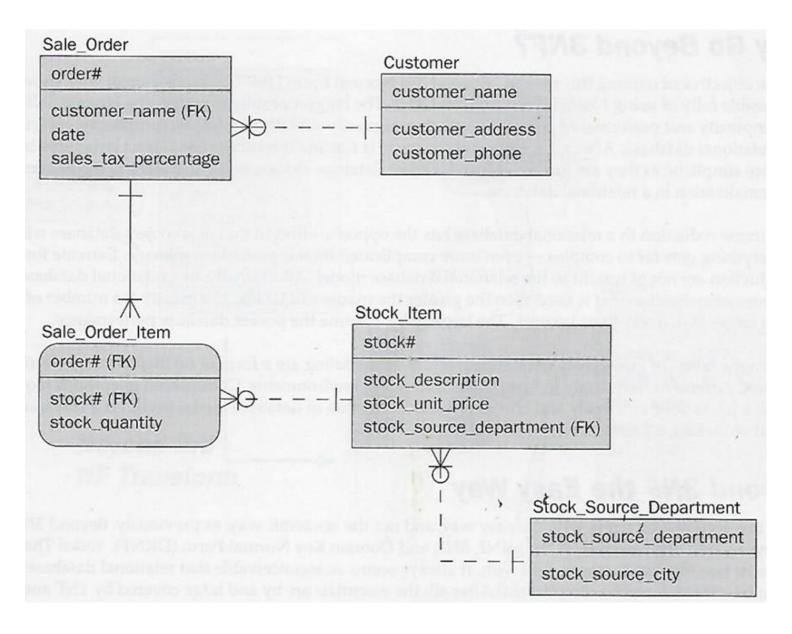
What foreign keys exist in the final database structure?

What referential integrity constraints existing in the final database structure?



What queries can we support with this

database?



Agenda

- ✓ Case Study Database design with input from legacy information systems
- ✓ Database Design Process / Data Normalization
- Reading a Database Design
- Database Design Security Considerations
- Database Design Assignment Review
- Quiz

Exercise:

Reading a database design

Working individually or in teams:

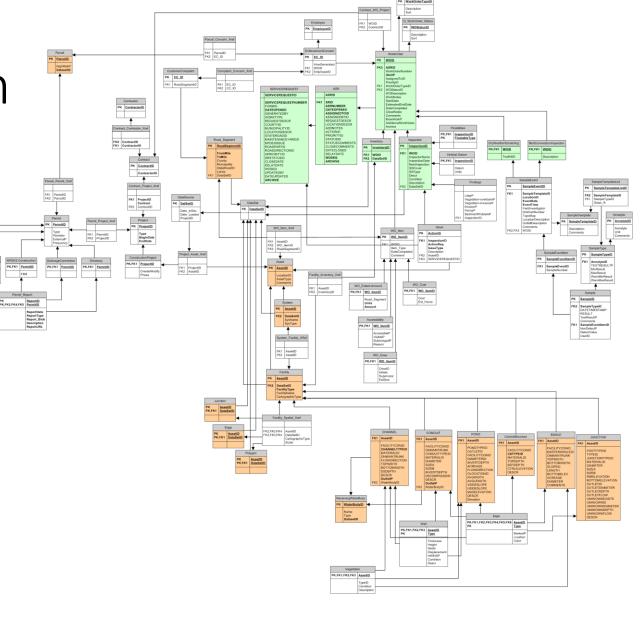
Access the following files available on the MIS Community Site's Wrap Up post for this class:

- Sr-asr-wo inventory location dpl 171304A.pdf
- Sr_asr-wo_inventory_location_dpl_071304.vsdx

Analyze the database design model and see if you can determine:

- i. What kind of information system is this?
 - i. What are the important data tables?
 - ii. What are the important relationships?
 - iii. What are the important attributes?
- ii. What can you figure out about
 - i. How this information system works?
 - ii. Who the users of this information system are?

Prepare a presentation of your findings When you are called on present your findings to the class



What kind of information system is this?

i. What are the important data tables?

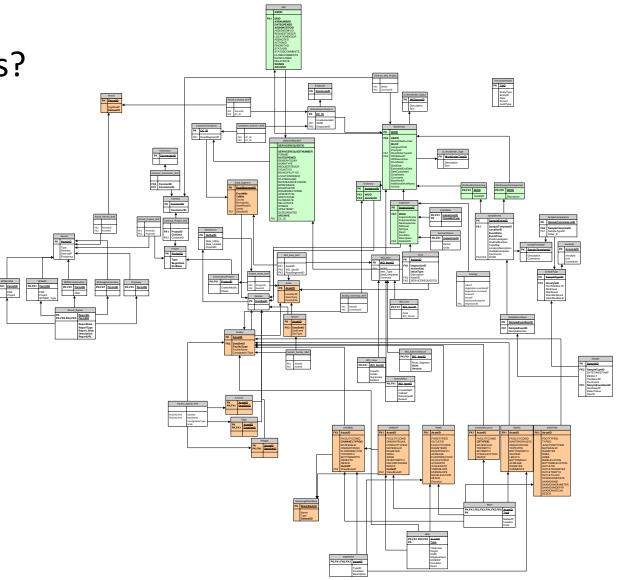
ii. What are the important relationships?

iii. What are the important attributes?

What can you figure out about

i. How this information system works?

ii. Who the users of this information system are?



Agenda

- ✓ Case Study Database design with input from legacy information systems
- ✓ Database Design Process / Data Normalization
- ✓ Reading a Database Design
- Database Design Security Considerations
- Database Design Assignment Review
- Quiz

Database Design Security considerations

- Almost all organizations maintain one more databases that contain information critical to operations,
 - e.g. customer information, human resources and benefits information, trade secrets...
- It is likely many contain personal information that users hold secret,
 - e.g. bank accounts, beneficiaries, credit card usage activities, expense reports, telephone records
- IT Auditors must help ensure that this information is identified, categorized, and protected by adequate security controls

General Data Protection Regulation (GDPR)

Created by the European Parliament, Council of the European Union, and European Commission to strengthen and unify data protection for all individuals within the European Union (EU)

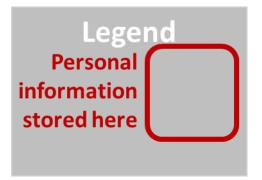


- Based on beliefs:
 - Privacy is a fundamental human right
 - Privacy protection is an essential element in the functioning of democratic societies
- Goals are:
 - 1. To give control back to citizens and residents over their personal data
 - 2. To simplify the regulatory environment for international businesses by unifying regulation within the EU
- Applies to organizations based outside the European Union collecting or processing personal data of EU residents, or exporting EU residents' personal data outside the EU

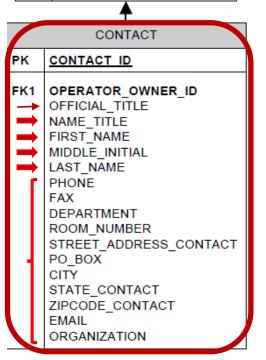
Personal data attributes can be classified based on disclosure potential:

- Identifiers. Attributes that unambiguously identify the subject (e.g. passport no., social security no., name-surname, etc.)
- Quasi-identifiers or key attributes. They identify the subject with some ambiguity, but their combination may lead to unambiguous identification (e.g. address, gender, age, tele-phone no., etc.)
- Confidential outcome attributes. They contain sensitive subject information (e.g. salary, religion, diagnosis, etc.)
- Non-confidential outcome attributes. Other attributes which contain non-sensitive subject entity information

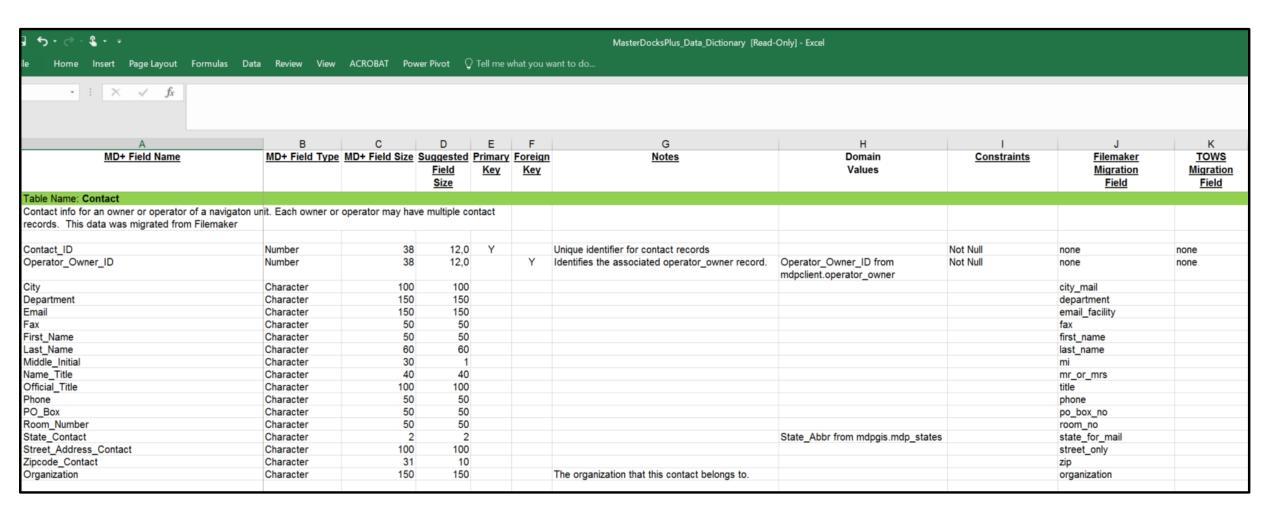
How to document where data subjects' personal information is stored and how it is used within a database?



	OPERATOR_OWNER
PK	OPERATOR OWNER ID
FK1	NAV_UNIT_GUID OPERATOR_OWNER_BOTH PUBLIC_OR_PRIVATE CURRENT_OR_PRIOR FACILITY_WEBSITE_URL NAME NAVUNIT_OR_VESSEL MASTER_OPERATOR_ID



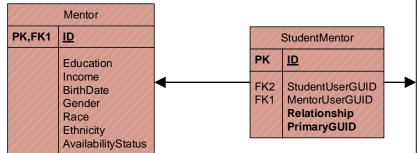
Personal data in Information System Database Data Dictionary

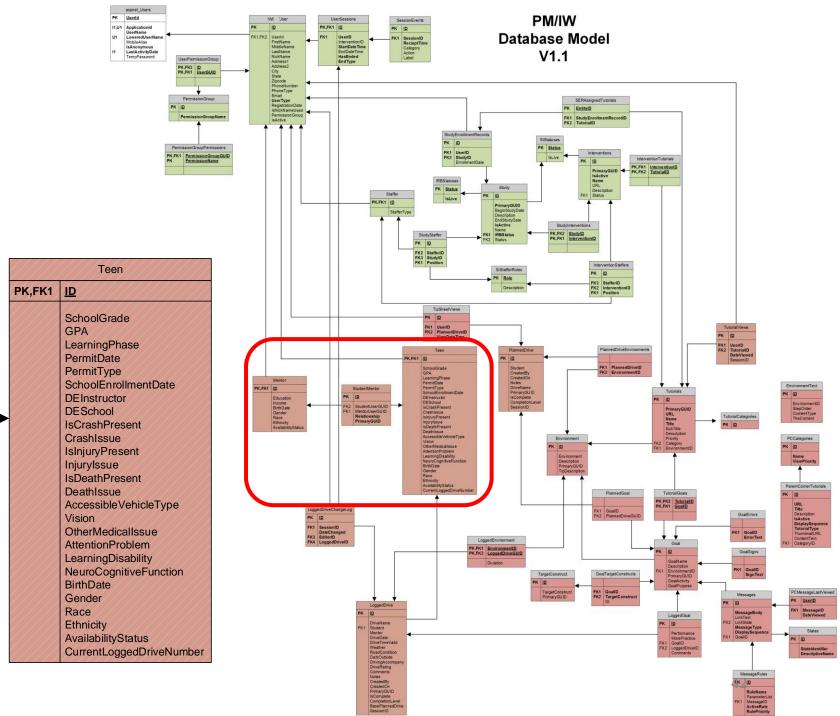


"Better" documentation identifies and classifies personal data in data dictionary

GDPR	MD+ Field Name	MD+ Field Type	MD+ Field Size		Primary	Foreign	<u>Notes</u>	Domain
Concern				<u>Field</u>	Key	Key		Values
Concern				<u>Size</u>				
	Table Name: Contact							
	Contact info for an owner or operator of a navigaton unit. Each owner or operator may have multiple contact records. This data was migrated from Filemaker							
Quasi-identifier	Contact_ID	Number	38	12,0			Unique identifier for contact records	
	Operator_Owner_ID	Number	38	12,0		Υ	Identifies the associated operator_owner	Operator_Owner_ID from
Quasi-identifier							record.	mdpclient.operator_owner
Quasi-identifier	City	Character	100	100				
Quasi-identifier	Department	Character	150	150				
Identifier	Email	Character	150	150				
Quasi-identifier	Fax	Character	50	50				
Identifier	First_Name	Character	50	50				
Identifier	Last_Name	Character	60	60				
Identifier	Middle_Initial	Character	30	1				
Quasi-identifier	Name_Title	Character	40	40				
Quasi-identifier	Official_Title	Character	100	100				
Quasi-identifier	Phone	Character	50	50				
Quasi-identifier	PO_Box	Character	50	50				
Quasi-identifier	Room_Number	Character	50	50				
Quasi-identifier	State_Contact	Character	2	2				State_Abbr from mdpgis.mdp_states
Quasi-identifier	Street_Address_Contact	Character	100	100				
Quasi-identifier	Zipcode_Contact	Character	31	10				
Quasi-identifier	Organization	Character	150	150			The organization that this contact belongs to.	

Is personal data identified in the database documentation?





Can you determine if data subjects' personal data is included in the data dictionary?

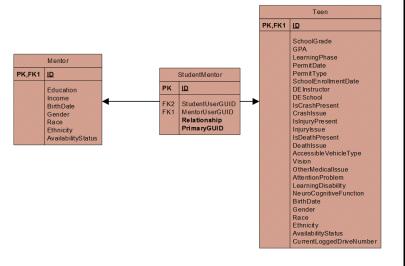
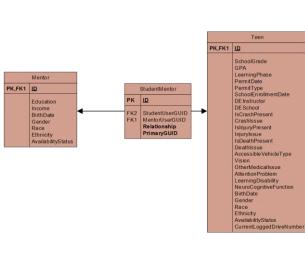


Table	Attribute	Primary Key	Foreign Key	Description
StudentMentor	StudentUserGUID		Yes	The student that is related to a mentor
StudentMentor	MentorUserGUID		Yes	The mentor that is related to a student
StudentMentor	Relationship			A description of the relationship of the mentor to the student
StudentMentor	ID	Yes		The ID of the relationship between the student and the mentor
StudentMentor	PrimaryGUID			A global unique identifier for the relationship
Mentor	ID	Yes	Yes	The ID of the mentor (and FK to USER)
Mentor	Education			The mentor's level of education
Mentor	Income			The mentor's level of income
Mentor	BirthDate			The birthdate of the mentor
Mentor	Gender			The mentor's gender
Mentor	Race			The mentor's race
Mentor	Ethnicity			The ethnicity of the mentor
Teen	ID	Yes	Yes	The ID of the teen/student - also the FK into the users table
Teen	SchoolGrade			The numeric school grade level at the time of enrollment
Teen	GPA			The decimal GPA of the teen at the time of enrollment
Teen	LearningPhase			Not implemented
Teen	PermitDate			The date that the teen's driving permit was issued
Teen	PermitType			The type of driving permit the teen has
Teen	SchoolEnrollmentDate			The date that the teen enrolled in driving school
Teen	DEInstructor			The teen's driving education instructor
Teen	DESchool			The name of the teen's driving school
Teen	IsCrashPresent			Indicates if the teen has a crash issue
Teen	CrashIssue			The nature of the crash issue
Teen	IsInjuryPresent			Indicates if an injury issue is present
Teen	Injurylssue			The nature of the injury issue
Teen	IsDeathPresent			Indicates if a death issue is present
Teen	DeathIssue			The nature of the death issue
Teen	AccessibleVehicleType			Teens access to a vehicle
Teen	Vision			The nature of any vision related issues
Teen	OtherMedicalIssue			The nature of any unclassified medical issues
Teen	AttentionProblem			Description of any attention defecit related issues
Teen	LearningDisability			Description of any learning disabilities
Teen	NeuroCognitiveFunction			Description of any neurocognitive function issues
Teen	BirthDate			The teen's birthdate
Teen	Gender			The teen's physical gender
Teen	Race			The race the teen most closely fits
Teen	Ethnicity			Indicates the teen's ethnicity - hispanic or not 43
Teen	CurrentLoggedDriveNumber			The number of logged drives associated with this teen as the driver

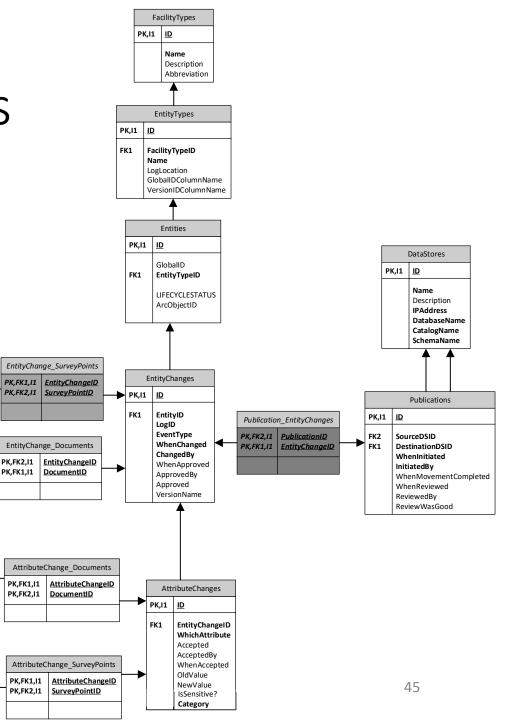
An improvement in identifying personal data and their classification in the data dictionary

GDPR



	GDPR					
	Concern	Table	Attribute	Primary Key	Foreign Key	Description
Qua	asi-identifier	StudentMentor	StudentUserGUID		Yes	The student that is related to a mentor
Qua	asi-identifier	StudentMentor	MentorUserGUID		Yes	The mentor that is related to a student
Con	fidential	StudentMentor	Relationship			A description of the relationship of the mentor to the student
Qua	si-identifier	StudentMentor	ID	Yes		The ID of the relationship between the student and the mentor
Qua	asi-identifier	StudentMentor	PrimaryGUID			A global unique identifier for the relationship
Qua	asi-identifier	Mentor	ID	Yes	Yes	The ID of the mentor (and FK to USER)
Cor	fidential	Mentor	Education			The mentor's level of education
Cor	fidential	Mentor	Income			The mentor's level of income
Cor	fidential	Mentor	BirthDate			The birthdate of the mentor
Cor	fidential	Mentor	Gender			The mentor's gender
Cor	fidential	Mentor	Race			The mentor's race
Cor	fidential	Mentor	Ethnicity			The ethnicity of the mentor
Qua	asi-identifier	Teen	ID	Yes	Yes	The ID of the teen/student - also the FK into the users table
Cor	fidential	Teen	SchoolGrade			The numeric school grade level at the time of enrollment
Cor	fidential	Teen	GPA			The decimal GPA of the teen at the time of enrollment
Nor	n-confidential	Teen	LearningPhase			Not implemented
Qua	asi-identifier	Teen	PermitDate			The date that the teen's driving permit was issued
Nor	n-confidential	Teen	PermitType			The type of driving permit the teen has
Qua	asi-identifier	Teen	SchoolEnrollmentDate			The date that the teen enrolled in driving school
Cor	fidential	Teen	DEInstructor			The teen's driving education instructor
Qua	asi-identifier	Teen	DESchool			The name of the teen's driving school
Cor	fidential	Teen	IsCrashPresent			Indicates if the teen has a crash issue
Cor	fidential	Teen	CrashIssue			The nature of the crash issue
Cor	fidential	Teen	IsInjuryPresent			Indicates if an injury issue is present
Cor	fidential	Teen	Injurylssue			The nature of the injury issue
Cor	fidential	Teen	IsDeathPresent			Indicates if a death issue is present
Cor	fidential	Teen	DeathIssue			The nature of the death issue
Cor	fidential	Teen	AccessibleVehicleType			Teens access to a vehicle
Cor	fidential	Teen	Vision			The nature of any vision related issues
Cor	fidential	Teen	OtherMedicalIssue			The nature of any unclassified medical issues
Cor	fidential	Teen	AttentionProblem			Description of any attention defecit related issues
Cor	fidential	Teen	LearningDisability			Description of any learning disabilities
Cor	fidential	Teen	NeuroCognitiveFunction			Description of any neurocognitive function issues
Qua	asi-identifier	Teen	BirthDate			The teen's birthdate
Qua	si-identifier	Teen	Gender			The teen's physical gender
Cor	fidential	Teen	Race			The race the teen most closely fits
Cor	fidential	Teen	Ethnicity			Indicates the teen's ethnicity - hispanic or not
Nor	n-confidential	Teen	CurrentLoggedDriveNumber			The number of logged drives associated with this teen as the driver

A database design for metadata capabilities for tracking changes to sensitive attribute values and their propagation through information system's data store as part of update processes



PK,FK2,I1

PK,FK1,I1

PK,FK2,I1

Documents

DocumentType

SurveyPoints DocumentDate

TEMPUMSID

DrawingDocumentName ImageDocumentName

SurveyDocumentName TEMPGID

PK,I1 ID

SurvevPoints

PointID

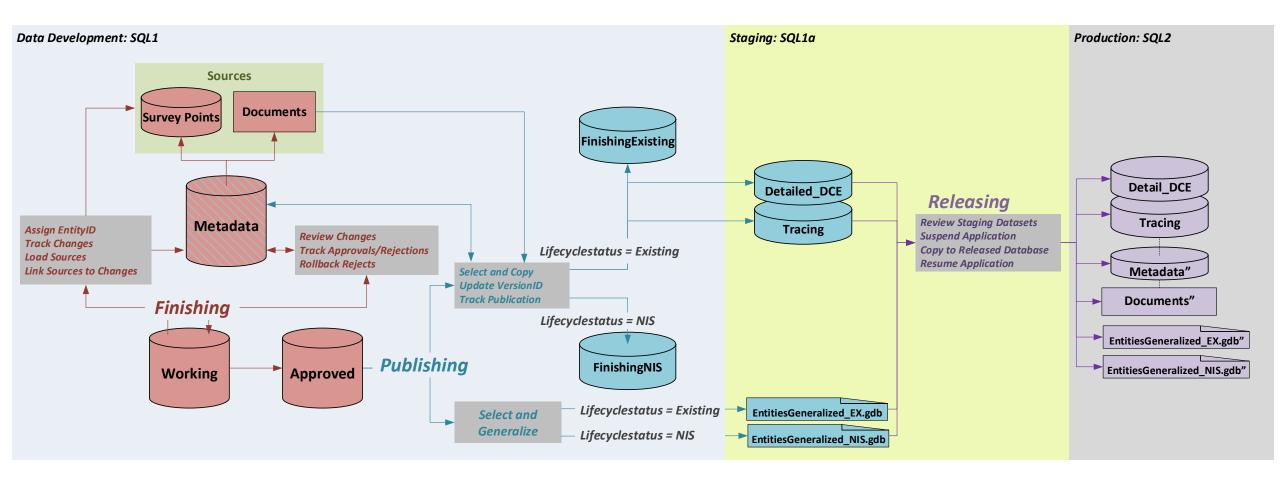
Northing

Easting

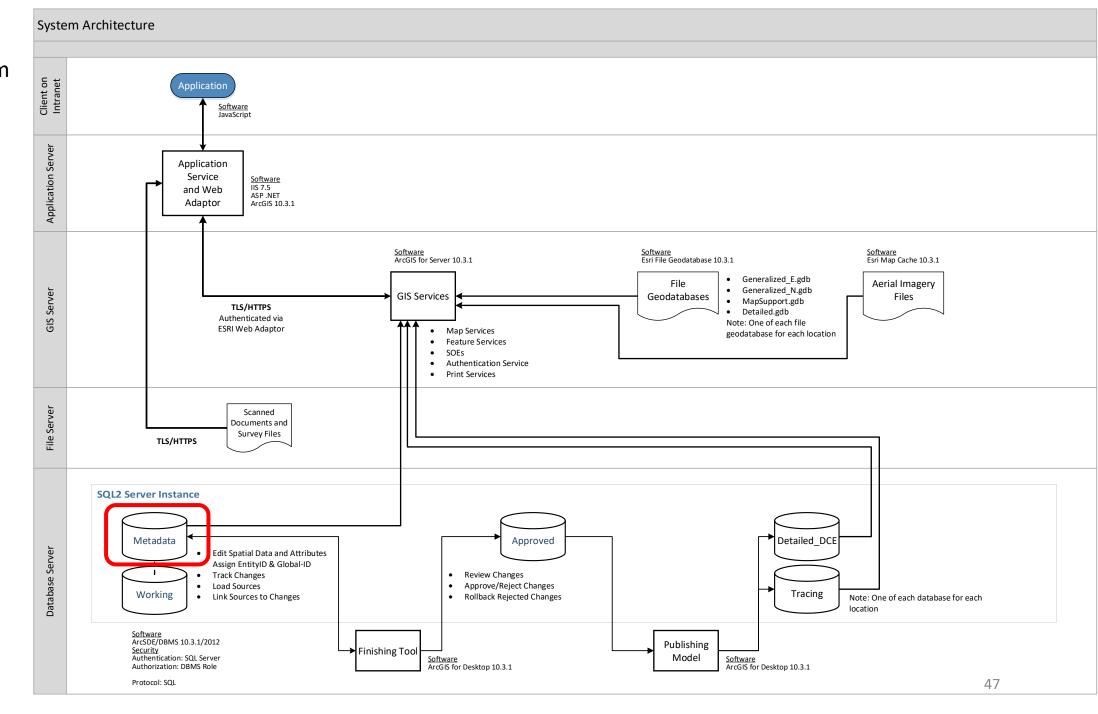
DocumentID

DescriptiveCode Surveyor

Example: Data update propagation documentation

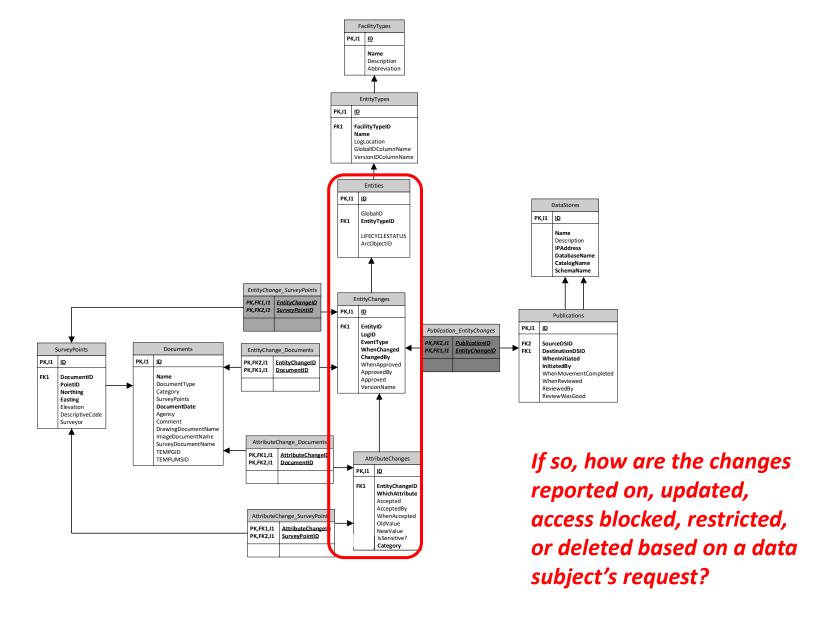


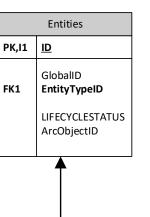
Production System Diagram



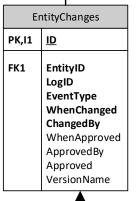
FacilityTypes Metadata support for PK,I1 ID Name Description Abbreviation tracking sources of EntityTypes PK,I1 ID FacilityTypeID attribute values and GlobalIDColumnName VersionIDColumnName their propagation Entities PK.I1 ID DataStores GlobalID PK.I1 ID FK1 EntityTypeID through a complex Name LIFECYCLESTATUS Description ArcObjectID **IPAddress** DatabaseName CatalogName SchemaName production EntityChange SurveyPoints EntityChanges PK,FK1,I1 | EntityChangeID PK,FK2,I1 | SurveyPointID ► PK,I1 ID Publications information system EntityID PK,I1 ID Publication EntityChanges LogID EventType FK2 SourceDSID PK.FK2.I1 Documents WhenChanged EntityChange Documents PK,FK1,I1 EntityChangeID DestinationDSID ChangedBy WhenInitiated PK,I1 ID **EntityChangeID** WhenApproved InitiatedBy PK.FK1.I1 **DocumentID** ApprovedBy WhenMovementCompleted Name DocumentID Approved WhenReviewed PointID DocumentType VersionName ReviewedBy Category Northing ReviewWasGood SurveyPoints Easting DocumentDate Elevation DescriptiveCode Agency Comment Surveyor DrawingDocumentName ImageDocumentName AttributeChange Documents SurveyDocumentName TEMPGID PK,FK1,I1 AttributeChangeID AttributeChanges **TEMPUMSID** PK,FK2,I1 DocumentID **EntityChangeID** WhichAttribute Accepted AcceptedBy AttributeChange SurveyPoints WhenAccepted OldValue AttributeChangeID NewValue 48 PK,FK2,I1 SurveyPointID IsSensitive? Category

Assessment questions

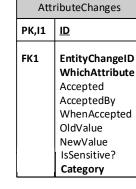




Which entities (if any) are data subject entities?

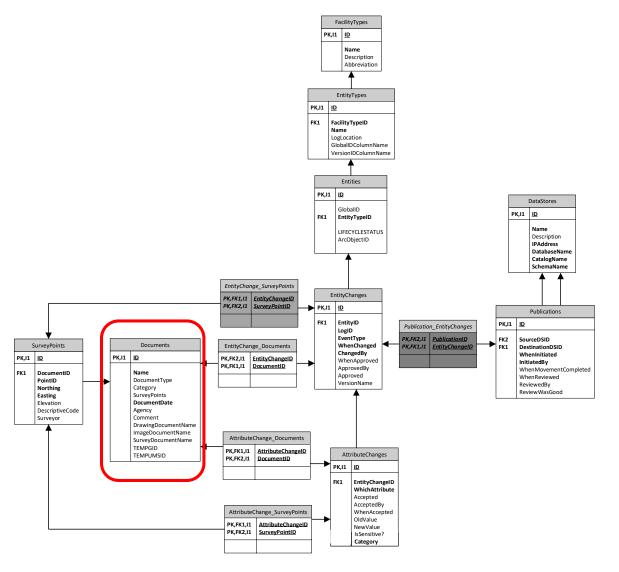


Do changes pertain to a data subject's personal data records?

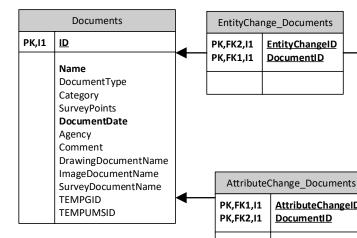


Do the changes pertain to a data subject's personal data attribute values? If so, which ones?

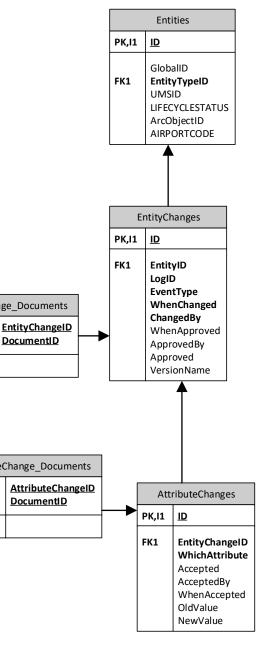
GDPR assessment questions



Do the documents contain sensitive personal data records?



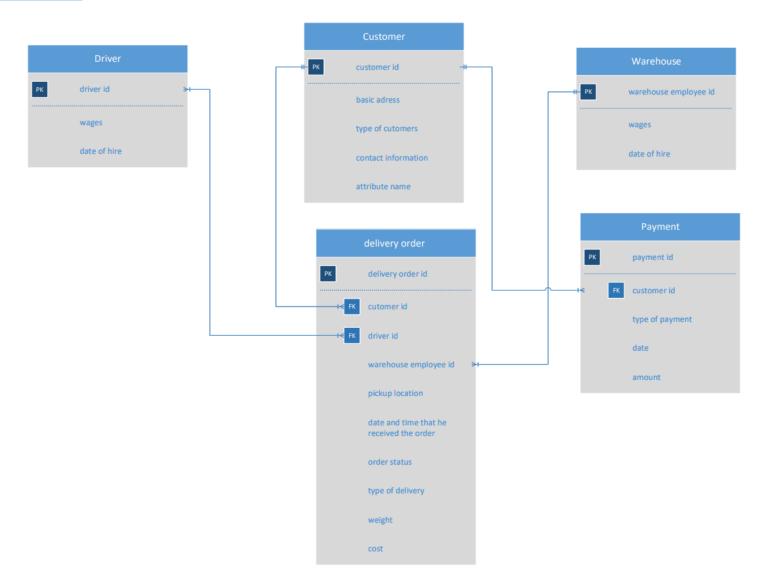
If so, how are they reported on, updated, access blocked, restricted, or deleted based on a data subject's request?

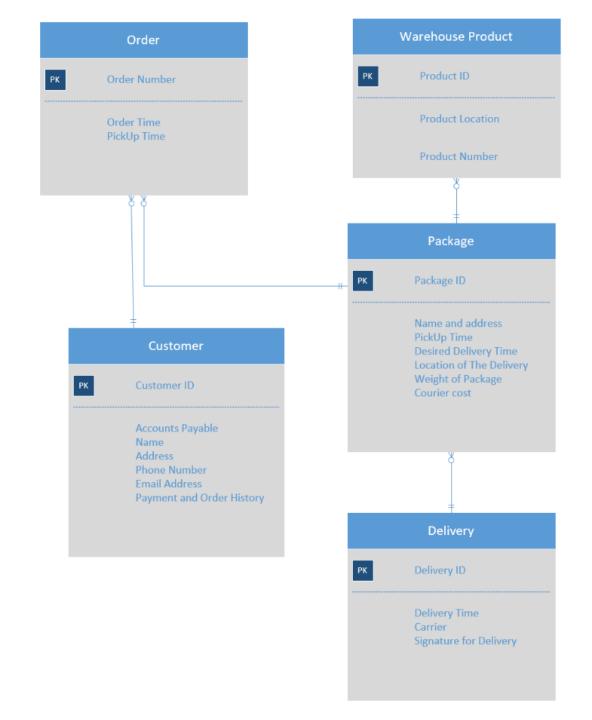


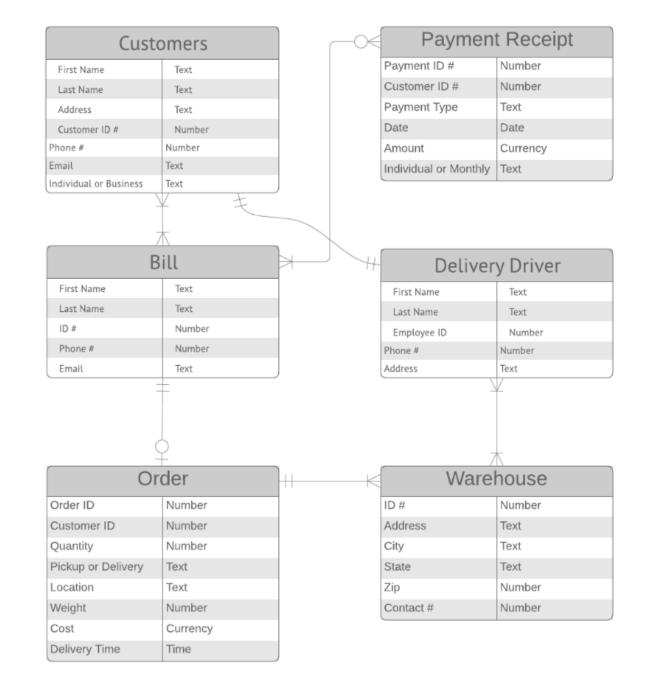
DocumentID

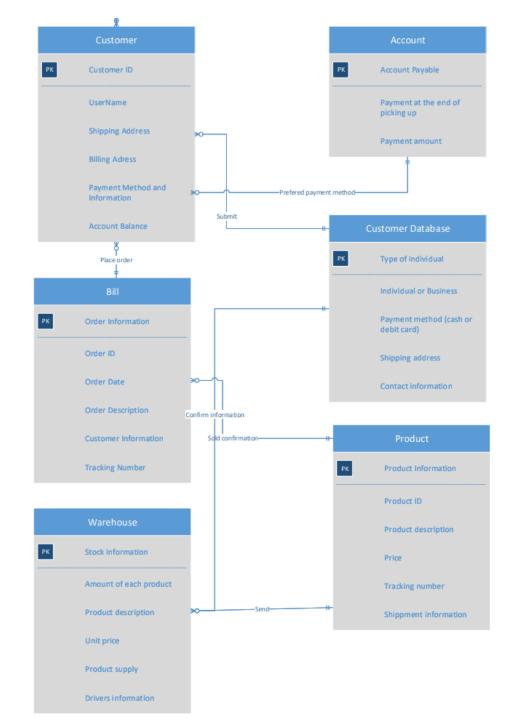
Agenda

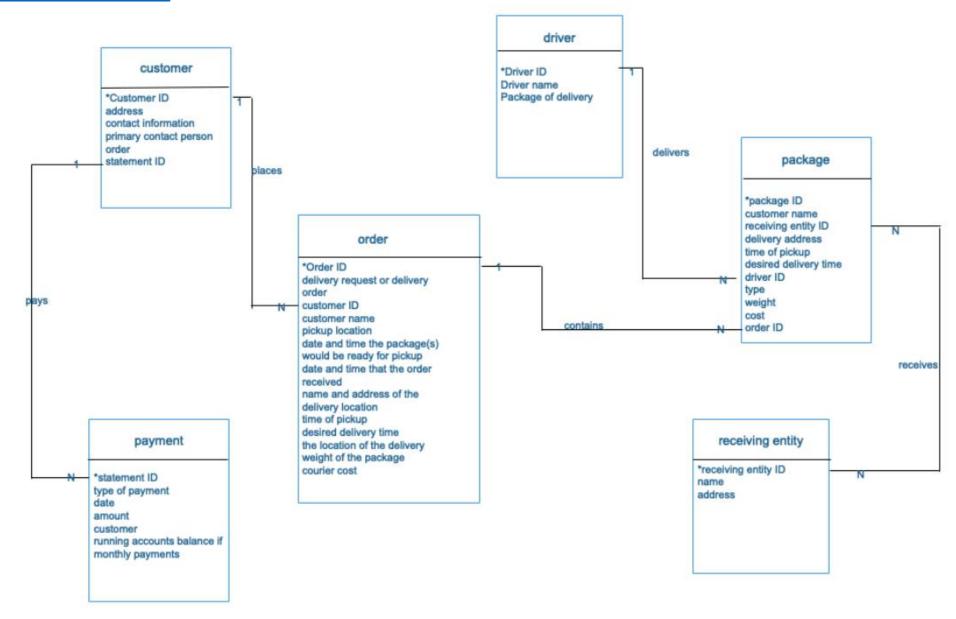
- ✓ Case Study Database design with input from legacy information systems
- ✓ Database Design Process / Data Normalization
- ✓ Reading a Database Design
- ✓ Database Design Security Considerations
- Database Design Assignment Review
- Quiz

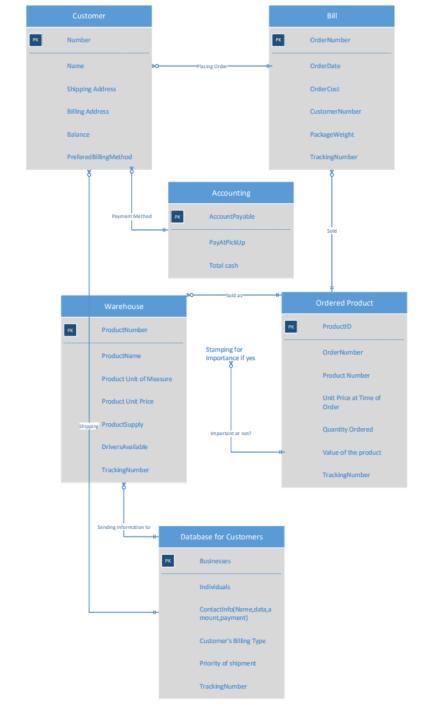


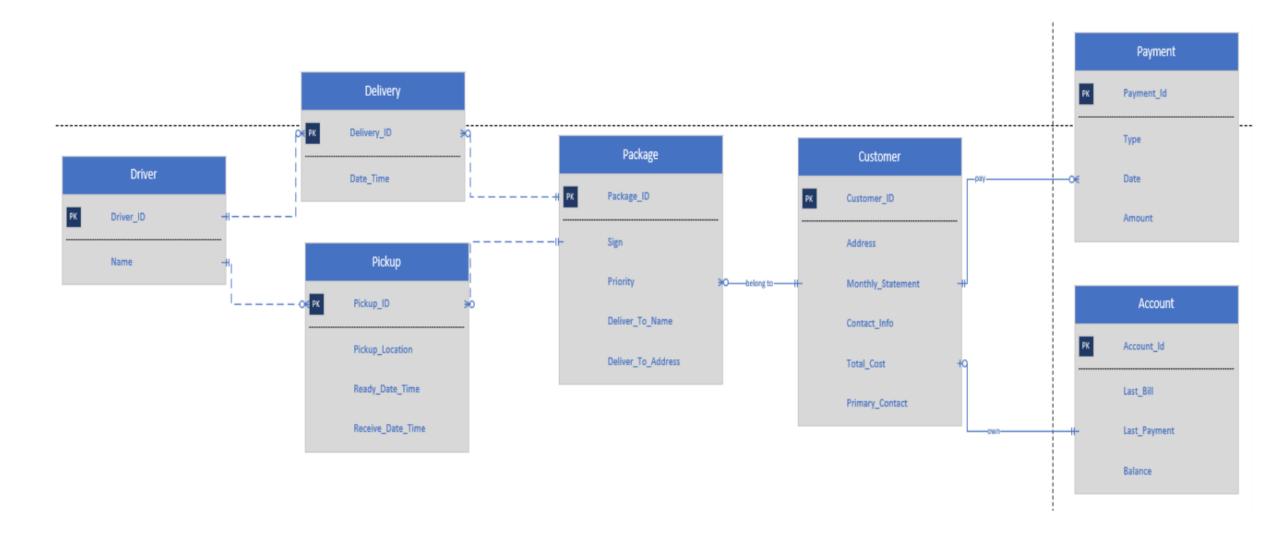


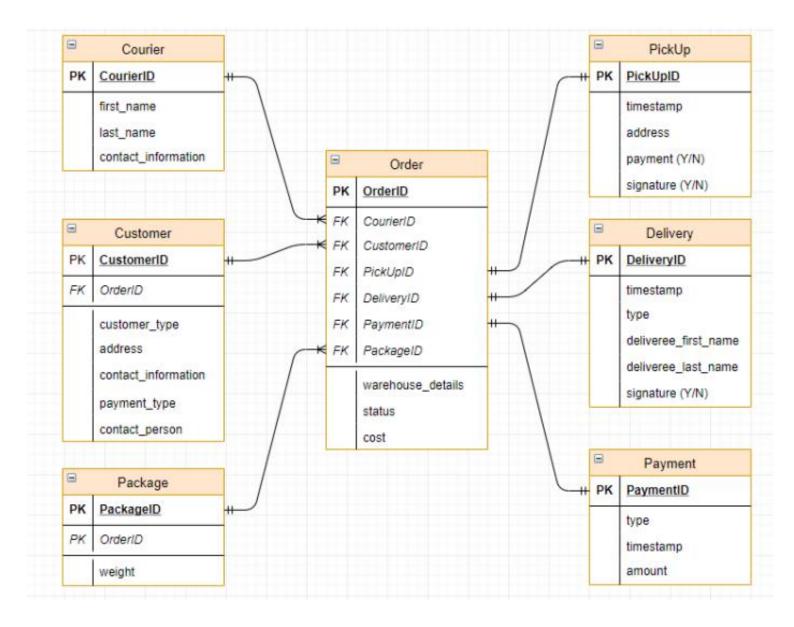


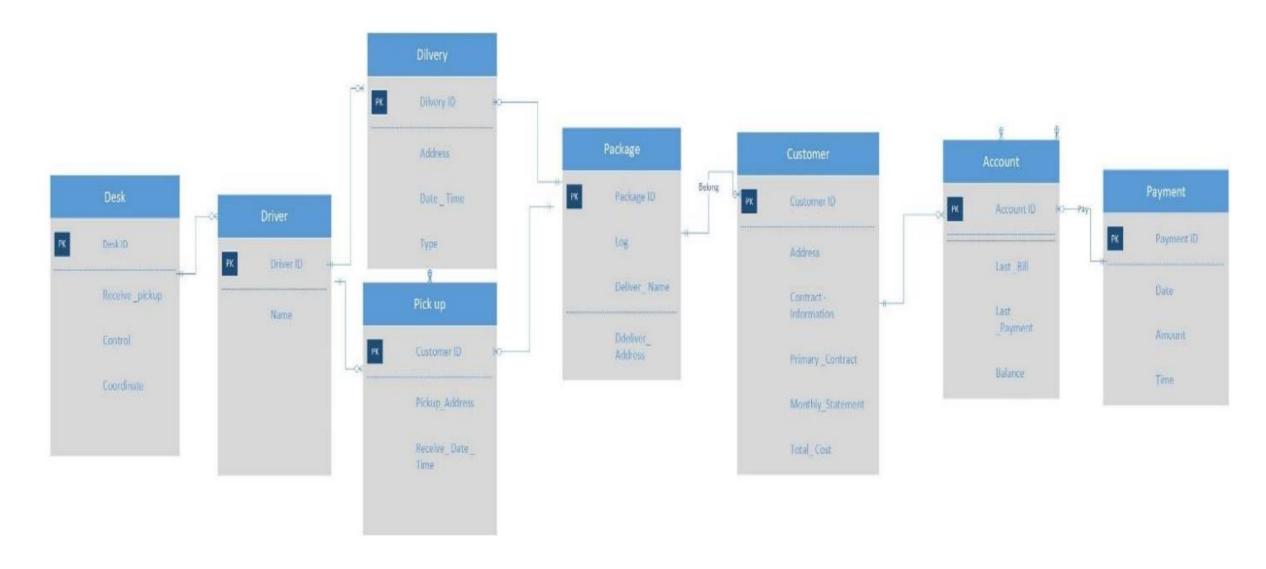


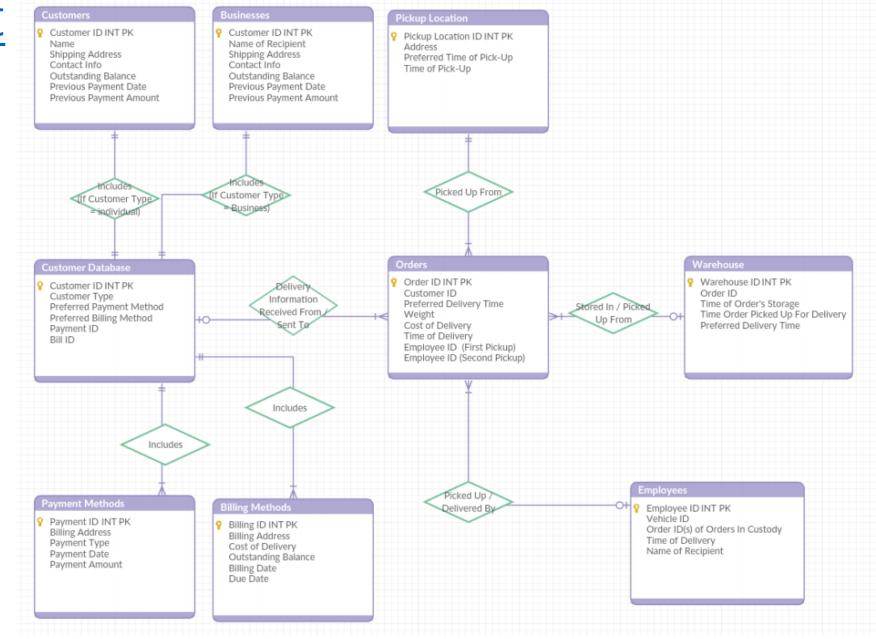


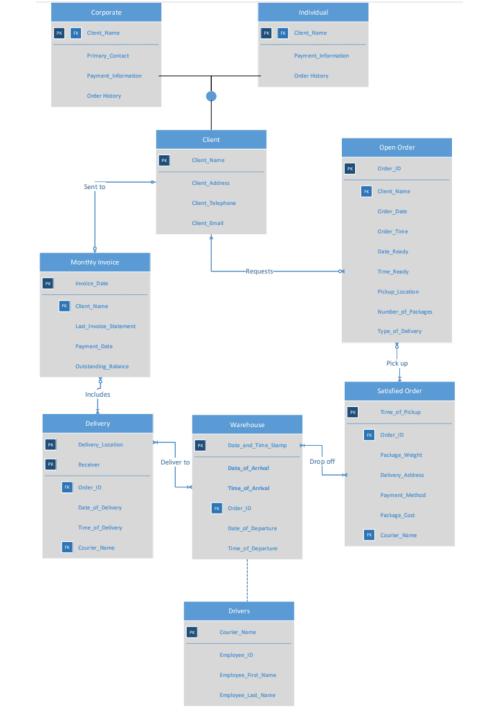


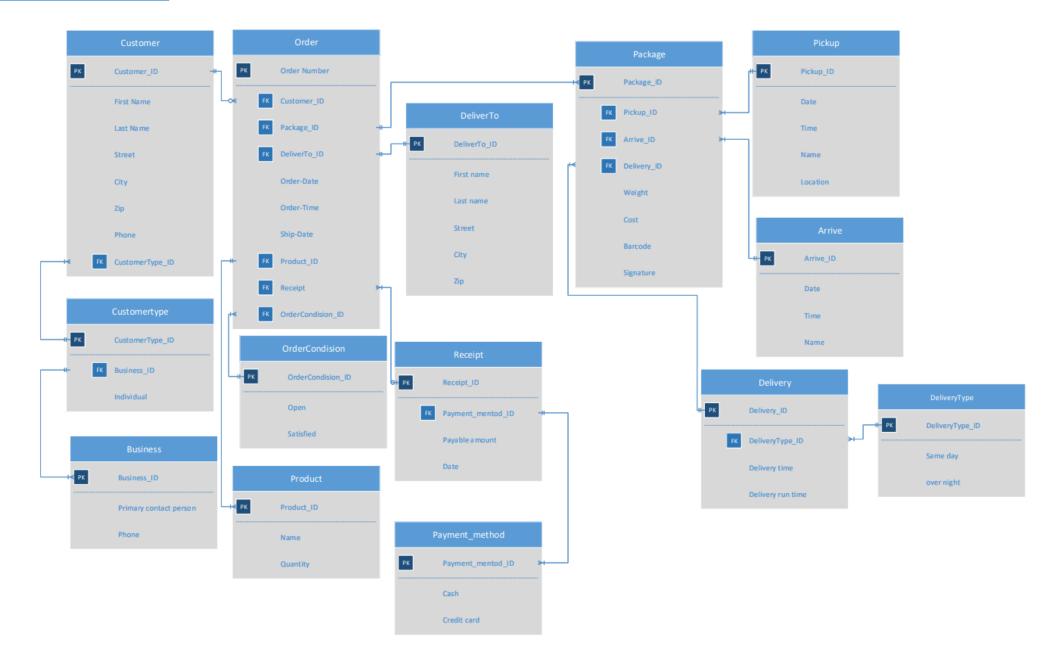


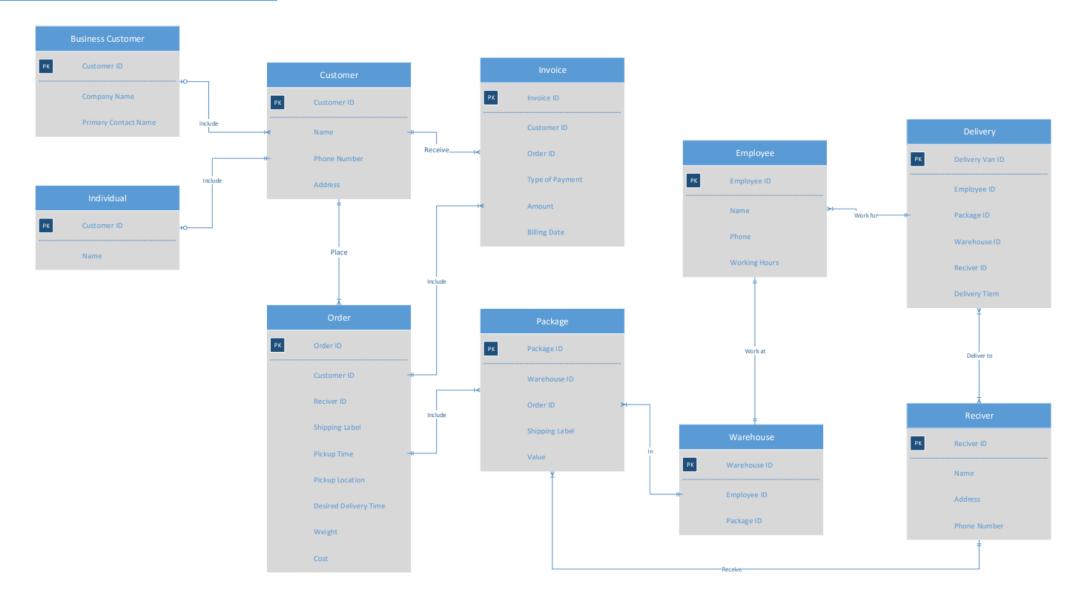


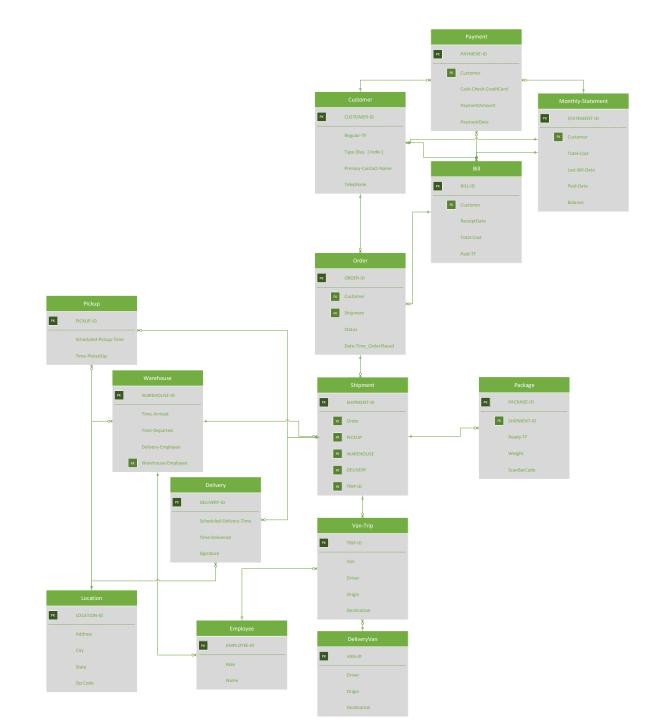












Quiz

Which one of the following database key types is used to enforce referential integrity relationships between tables?

- a. Primary key
- b. Candidate key
- c. Foreign key
- d. Master key

Which of the following database keys is used by a DBMS to uniquely identify each row in a database table?

- a. Foreign key
- b. Primary key
- c. Candidate key
- d. Referential key

Agenda

- ✓ Case Study Database design with input from legacy information systems
- ✓ Database Design Process / Data Normalization
- ✓ Reading a Database Design
- ✓ Database Design Security Considerations
- ✓ Database Design Assignment Review
- **√** Quiz