DATA WAREHOUSING
LECTURE 4

ENGR. MADEHA MUSHTAQ
DEPARTMENT OF COMPUTER SCIENCE
IQRA NATIONAL UNIVERSITY

META DATA

- Metadata in a data warehouse contains the answers to questions about the data in the data warehouse.
- We keep the answers in a place called the metadata repository.
- Here is a sample list of definitions of Metadata:

META DATA

- Data about the data
- Table of contents for the data
- Catalog for the data
- Data warehouse roadmap
- Data warehouse directory
- Glue that holds the data warehouse contents together

WHO NEEDS METADATA?

- Imagine a filing cabinet stuffed with documents without any folders and labels.
- Without metadata, your data warehouse is like such a filing cabinet.
- It is probably filled with information very useful for your users and for IT developers and administrators.
- But without any easy means to know what is there, the data warehouse is of very limited value.

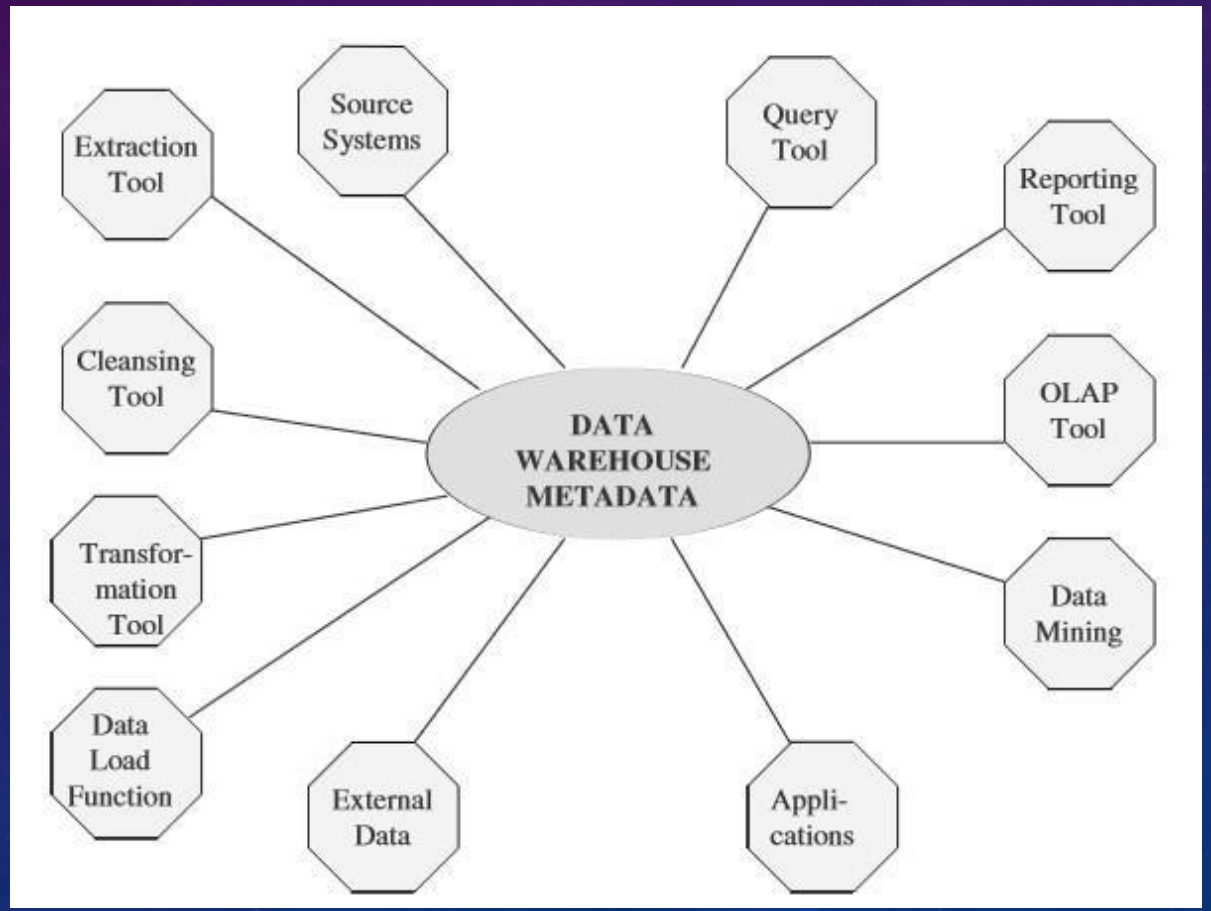
WHO NEEDS METADATA?

	IT Professionals	Power Users	Casual Users
Information Discovery	Databases, Tables, Columns, Server Platforms	Databases, Tables, Columns	List of Predefined Queries and Reports, Business Views
Meaning of Data	Data Structures, Data Definitions, Data Mapping, Cleansing Functions, Transformation Rules	Business Terms, Data Definitions, Data Mapping, Cleansing Functions, Transformation Rules	Business Terms, Data Definitions, Filters, Data Sources, Conversion , Data Owners
Information Access	Program Code in SQL, 3GL,4GL, Front-end Applications, Security	Query Toolsets, Database Access for Complex Analysis	Authorization Requests, Information Retrieval into Desktop Applications such as Spreadsheets

METADATA IS LIKE A NERVE CENTER

- Various processes during the building and administering of the data warehouse generate parts of the data warehouse metadata.
- Parts of metadata generated by one process are used by another.
- In the data warehouse, metadata assumes a key position and enables communication among various processes. It acts like a nerve center in the data warehouse.

METADATA IS LIKE A NERVE CENTER



WHY METADATA IS VITAL FOR END-USERS

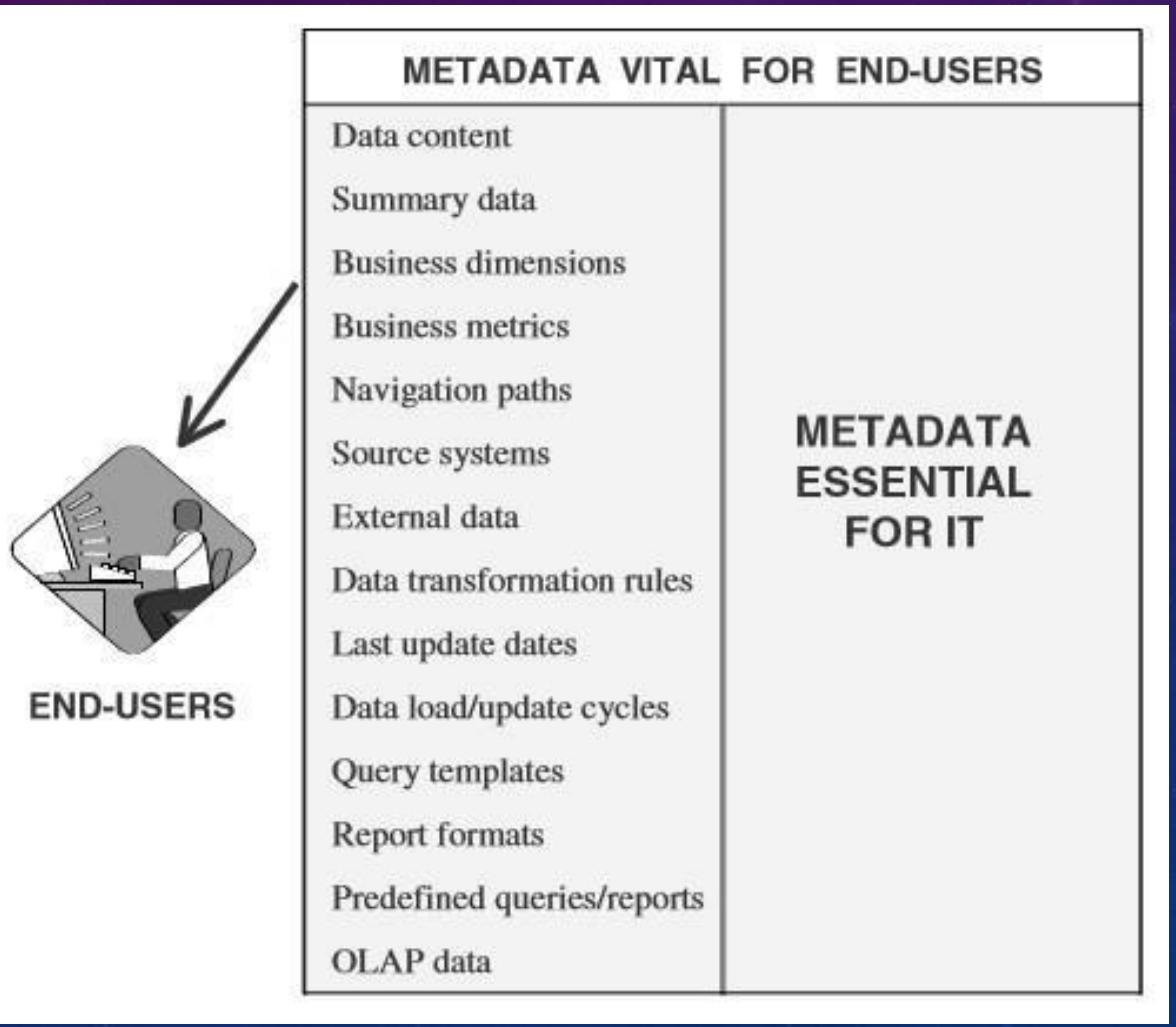


Figure shows the types of information metadata provides to the end-users and the purposes for which they need these types of information.

WHY METADATA IS ESSENTIAL FOR IT

For performing the responsibilities for design and administration, IT must have access to proper metadata.

METADATA ESSENTIAL FOR IT	
METADATA VITAL FOR END-USERS	Source data structures Source platforms Data extraction methods External data Data transformation rules Data cleansing rules Staging area structures Dimensional models Initial loads Incremental loads Data summarization OLAP system Web-enabling Query/report design

IT PROFESSIONALS

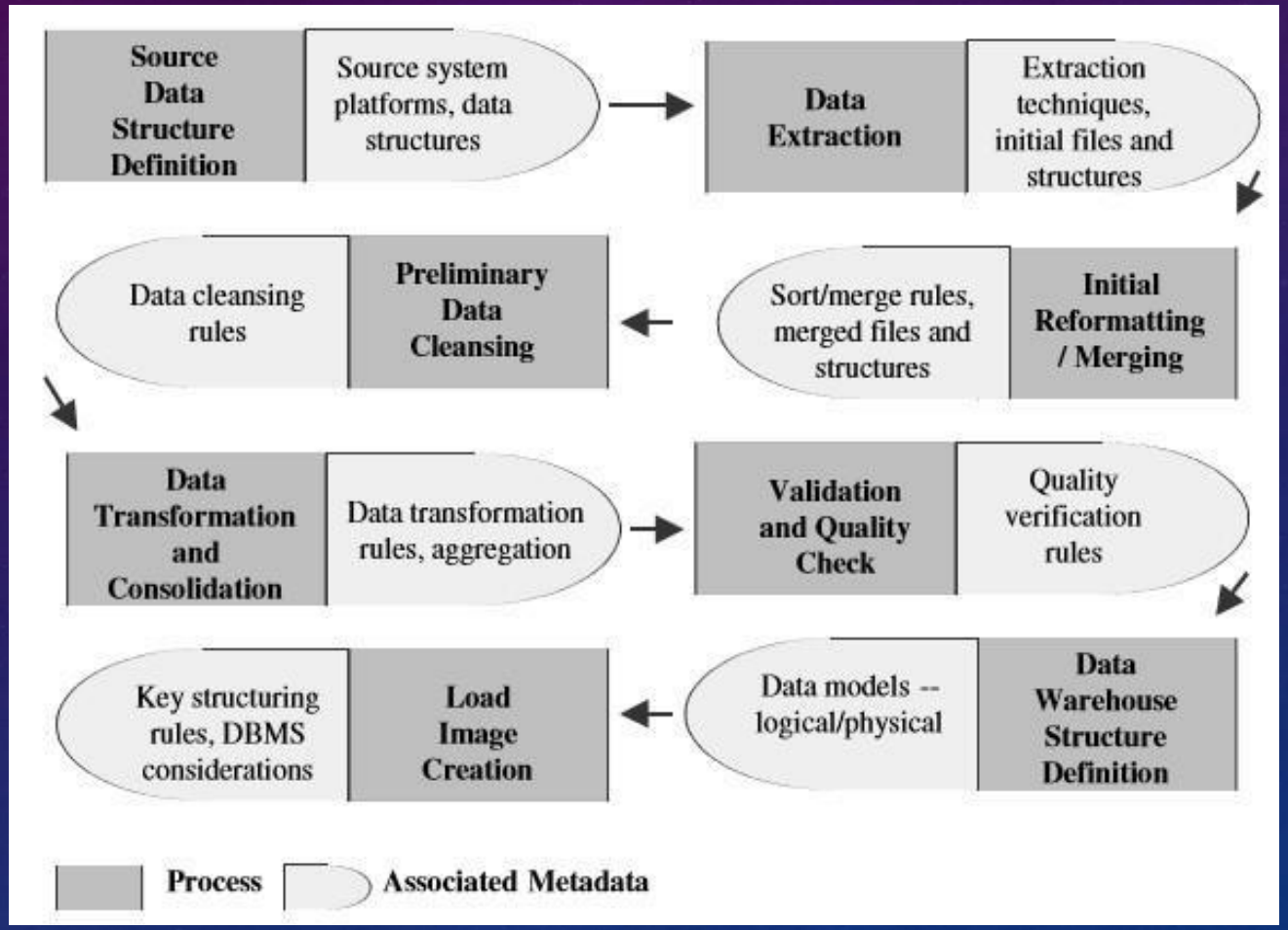
AUTOMATION OF WAREHOUSING TASKS

- As the data movement takes place from the data sources to the data warehouse database through the data staging area, several processes occur.
- In a typical data warehouse, appropriate tools assist in these processes.
- Each tool records its own metadata as data movement takes place.
- The metadata recorded by one tool drives one or more processes that follow. This is how metadata assumes an active role and assists in the automation of data warehouse processes.

AUTOMATION OF WAREHOUSING TASKS

- Here is a list of back-end processes shown in the order in which they generally occur:
- Source data structure definition
- Data extraction
- Initial reformatting/merging
- Preliminary data cleansing
- Data transformation and consolidation
- Validation and quality check
- Data warehouse structure definition
- Load image creation

AUTOMATION OF WAREHOUSING TASKS



Metadata drives data warehouse processes

METADATA TYPES BY FUNCTIONAL AREAS

- Classification of metadata types by the functional areas in the data warehouse:
- Data acquisition
- Data storage
- Information delivery

DATA ACQUISITION

- In this area, the data warehouse processes relate to the following functions:
- Data extraction
- Data transformation
- Data cleansing
- Data integration
- Data staging

DATA ACQUISITION

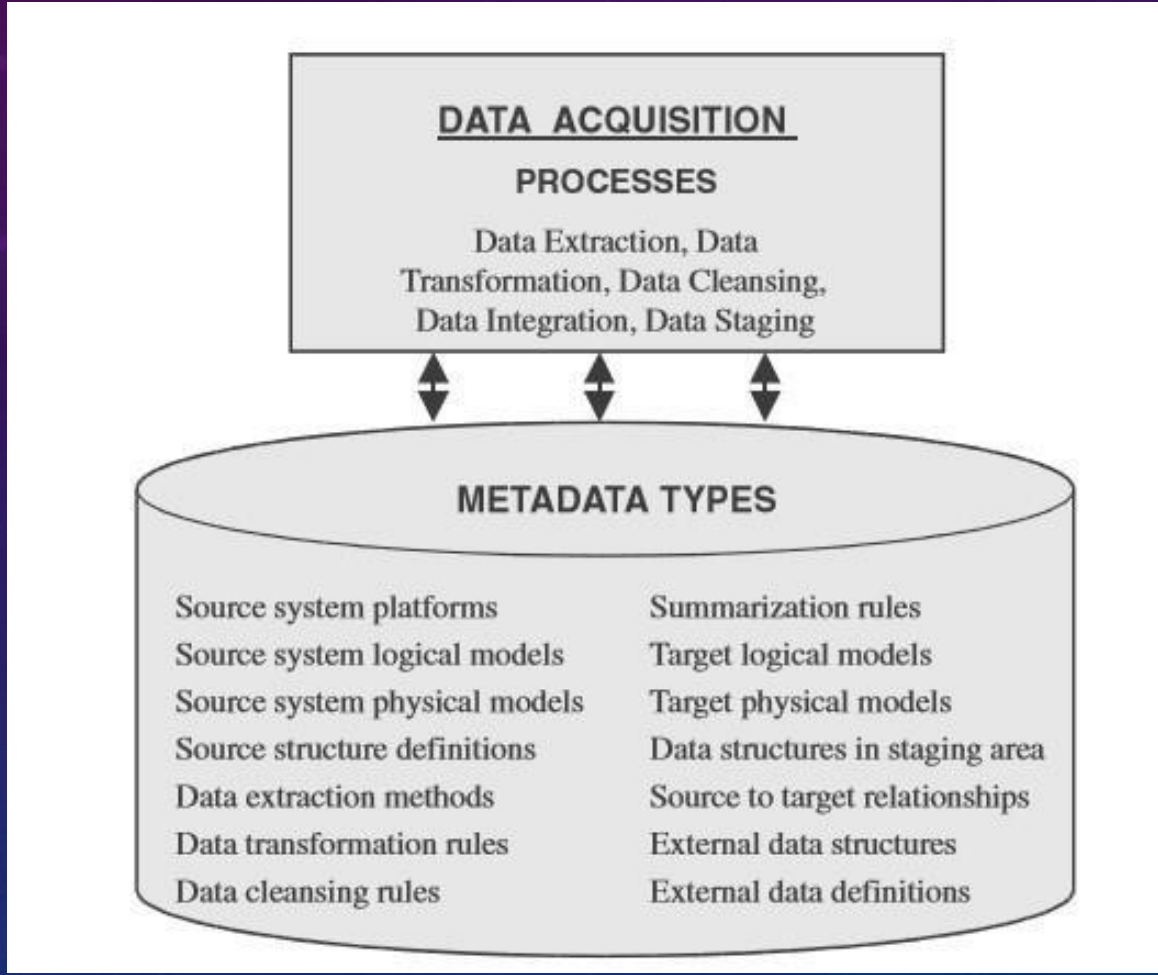


Figure shows metadata types recorded and used in the data acquisition area.

DATA STORAGE

- In this area, the data warehouse processes relate to the following functions:
- Data loading
- Data archiving
- Data management

DATA STORAGE

- Just as in the other areas, as processes take place in the data storage functional area, the appropriate tools record the metadata elements relating to the processes.
- Metadata recorded by processes in the data storage area is used for development, administration, and by the users.
- You will be using the metadata from this area for designing the full data refreshes and the incremental data loads.
- The DBA will be using metadata for the processes of backup and recovery.

DATA STORAGE

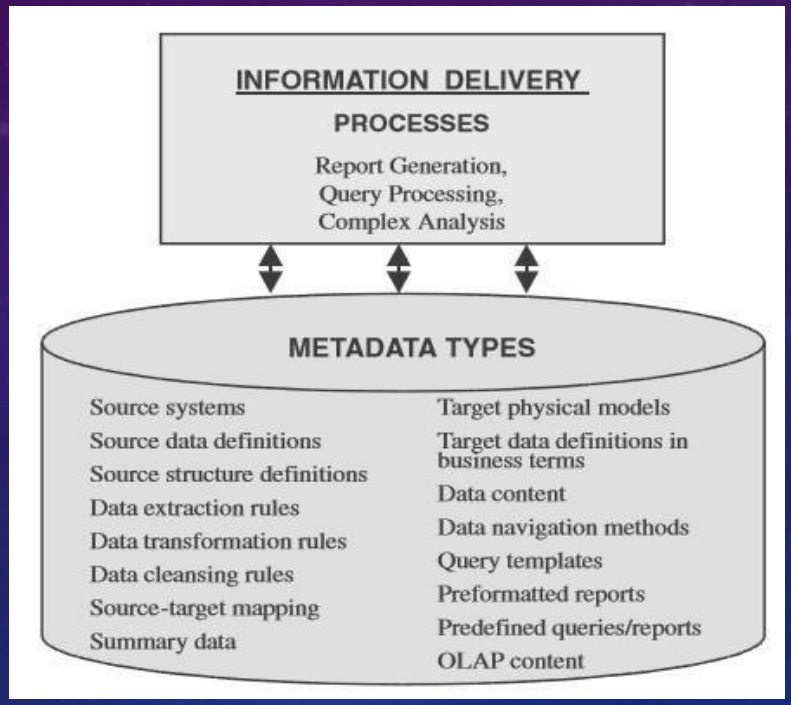


Figure shows metadata types recorded and used in the data storage area.

INFORMATION DELIVERY

- In this area, the data warehouse processes relate to the following functions:
- Report generation
- Query processing
- Complex analysis

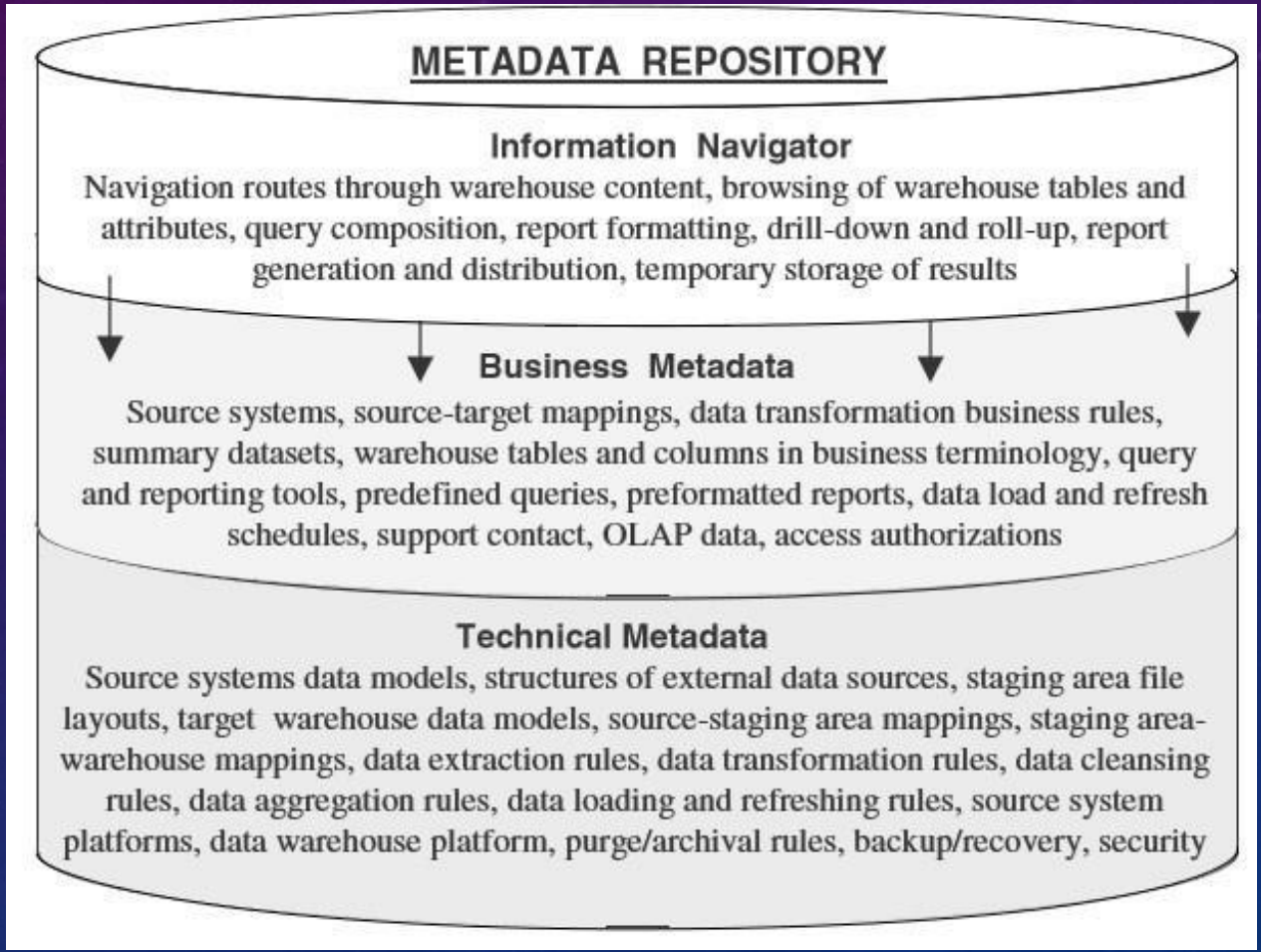
INFORMATION DELIVERY

- Mostly, the processes in this area are meant for end-users.
- While using the processes, end-users generally use metadata recorded in processes of the other two areas of data acquisition and data storage.
- The user can find the date of the last full refresh and the incremental loads for various tables in the data warehouse database.
- Generally, metadata recorded in the information delivery functional area relate to predefined queries, predefined reports, and input parameter definitions for queries and reports.

METADATA REPOSITORY

- Think of a metadata repository as a general-purpose information directory or cataloguing device to classify, store, and manage metadata.
- The metadata repository can be thought of as two distinct information directories, one to store business metadata and the other to store technical metadata.
- This division may also be logical within a single physical repository.

METADATA REPOSITORY



END OF SLIDES