CHAPTER 5: Clarifying the Research Question through Secondary Data and Exploration (Handout)

A SEARCH STRATEGY FOR EXPLORATION

- Exploration is particularly useful when researchers lack a clear idea of the problems they will meet during the study.
  - Exploration is the phase through which researchers develop concepts more clearly, establish priorities, develop operational definitions, and improve the final research design.
  - Exploration may save time and money.
  - Exploration is needed when studying new phenomena or situations.
  - Exploration is often, however, given less attention than it deserves.
- The exploratory phase search strategy usually comprises one or more of the following:
  - Discovery analysis of secondary sources such as published studies, document analysis, and retrieval of information from organizations' databases.
  - Interviews with those knowledgeable about the problem or its possible solutions (called **expert interviews**).
  - Interviews with individuals involved with the problem (called **individual depth interviews (IDIs)**—a type of interview that encourages the participant to talk extensively, sharing as much information as possible).
  - Group discussion with individuals involved with the problem or its possible solutions (including informal groups, as well as formal techniques such as focus groups or brainstorming).
  - A review of secondary sources is considered critical to moving from management question to research question.
  - The objective of the exploratory research phase:
    1. Expand your understanding of the management dilemma by looking for ways others have addressed and/or solved problems similar to your management dilemma or management question.
    2. Gather background information on your topic to refine the research question.
    3. Identify information that should be gathered to formulate investigative questions.
    4. Identify sources for and actual questions that might be used as measurement questions.
    5. Identify sources for and actual sample frames (lists of potential participants) that might be used in sample design.
  - In most cases, the exploration phase will begin with a **literature search**—a review of books, articles, research studies, or Web-published materials related to the proposed study.
  - In general, a literature search has five steps:
    1. Define your management dilemma or **management question**.
    2. Consult encyclopedias, dictionaries, handbooks, and textbooks to identify key terms, people, or events relevant to the management dilemma or management question.
    3. Apply these key terms, names of people, or events in searching indexes, bibliographies, and the Web to identify specific secondary sources.
    4. Locate and review specific secondary sources for relevance to your management dilemma.
    5. Evaluate the value of each source and its content.
  - Often the literature search leads to the research proposal.

- This proposal covers at minimum a statement of the research question and a brief description of the proposed research methodology.
- The proposal summarizes the findings of the exploratory phase of the research, usually with a bibliography of secondary sources that have led to the decision to propose a formal research study.

- Levels of Information: Information sources are generally categorized into three levels:
  - **Primary sources** are original works of research or raw data without interpretation or pronouncements that represent an official opinion or position.
    - Primary sources are always the most authoritative because the information has not been filtered or interpreted by a second party.
  - **Secondary sources** are interpretations of primary data.
    - Nearly all reference materials fall into this category.
    - A firm searching for secondary sources can search either internally or externally.
  - **Tertiary sources** are aids to discover primary or secondary sources or an interpretation of a secondary source.
    - These sources are generally represented by indexes, bibliographies, or Internet search engines.
- It is important to remember that all information is not of equal value.
  - Primary sources are the most valuable.

- Types of Information Sources.
  - Indexes and Bibliographies.
    - An **index** is a secondary data source that helps identify and locate a single book, journal article, author, et cetera, from among a large set.
    - A **bibliography** is an information source that helps locate a single book, article, photograph, et cetera.
    - Today, the most important bibliography in any library is its online catalog.
    - Skill in searching bibliographic databases is essential for any business researcher.
  - Dictionaries.
    - **Dictionaries** are secondary sources that define works, terms or jargon unique to a discipline; may include information on people, events, or organizations that shape the discipline; an excellent source of acronyms.
    - There are many specialized dictionaries that are field specific (e.g., business concepts dictionaries).
    - A growing number of dictionaries are found on the Web.
  - Encyclopedias.
    - An **encyclopedia** is a secondary source that provides background or historical information on a topic.
    - In addition to finding facts, encyclopedias are useful in identifying experts in a field or in finding key writings on any topic.
  - Handbooks.
    - A **handbook** is a secondary source used to identify key terms, people, or events relevant to the management dilemma or management question.
    - Handbooks often include statistics, directory information, a glossary of terms, and other data such as laws and regulations essential to a field.
    - The best handbooks include source references for the facts they present.

- One of the most important handbooks for business-to-business organizations is the North American Industry Classification System, United States (NAICS).
- Directories.
  - A **directory** is a reference source used to identify contact information.
  - Today, many directories are available at no charge via the Internet.
  - Most comprehensive directories are proprietary.

- Evaluating Information Sources.
  - Researchers should evaluate and select information sources based on five factors that can be applied to any type of source, these are:
    - *Purpose*—the explicit or hidden agenda of the information source.
    - *Scope*—the breadth and depth of topic coverage, including time period, geographic limitations, and the criteria for information inclusion.
    - *Authority*—the level of the data (primary, secondary, tertiary) and the credentials of the source author(s).
    - *Audience*—the characteristics and background of the people or groups for whom the source was created.
    - Format—how the information is presented and the degree of ease of locating specific information within the source.

MINING INTERNAL SOURCES

- The term **data mining** describes the process of discovering knowledge from databases stored in **data marts** or data warehouses.
  - The purpose of data mining is to identify valid, novel, useful, and ultimately understandable patterns in data.
  - Data mining is an approach that combines exploration and discovery with confirmatory analysis.
- An organization's own internal historical data is an often under-utilized source of information in the exploratory phase.
  - The researcher may lack knowledge that such historical data exist; or,
  - The researcher may choose to ignore such data due to time or budget constraints, and the lack of an organized archive.
  - Digging through data archives can be as simplistic as sorting through a file of patient records or inventory shipping manifests, or rereading company reports and management authored memos
- A **data warehouse** is an electronic repository for databases that organizes large volumes of data into categories, to facilitate retrieval, interpretation, and sorting by end users.
  - The data warehouse provides an accessible archive to support dynamic organizational intelligence applications.
  - The key words here are dynamically accessible. Data in a data warehouse must be continually updated to ensure that managers have access to data appropriate for real-time decisions.
  - The more accessible the databases that comprise the data warehouse, the more likely a researcher will use such databases to reveal patterns. Thus, researchers are more likely to mine electronic databases than paper ones.
- Remember that data in a data warehouse were once primary data, collected for a specific purpose.

- When researchers data-mine a company's data warehouse, all the data contained within that database have become secondary data.
- The patterns revealed will be used for purposes other than those originally intended.
  - Example: An archive of sales invoices provides a wealth of data about what was sold, at what price, and to whom, where, when, and how the products were shipped. But the company initially generated the sales invoice to facilitate the process of getting paid for the items shipped.
  - When a researcher mines the sales invoice archive, the search is for patterns of sales, by product, category, region, price, shipping methods, etc.
- Data mining forms a bridge between primary and secondary data.

Pattern Discovery

- Data-mining tools can be programmed to sweep regularly through databases and identify previously hidden patterns.
  - An example of pattern discovery is the detection of stolen credit cards based on analysis of credit card transaction records.
  - Other uses include:
    - Technical pattern in FOREX and stock markets.
    - Finding retail purchase patterns (used for inventory management)
    - Identifying call center volume fluctuations (used for staffing)
    - Locating anomalous data that could represent data entry errors (used to evaluate training, employee evaluation, or security needs)

Predicting Trends and Behaviors

- Data mining is used to predict problems:
  - A typical example of a predictive problem is targeted marketing.
  - Forecasting bankruptcy and loan default
  - Finding population segments with similar responses to a given stimulus.
  - Build risk models for a specific market, such as discovering the top 10 most significant buying trends each week.

Data-Mining Process

Sample

- The researcher must decide whether to use the entire data set (census) or a sample of the data.
  - If the data set in question is not large, if processing power is high, or if it is important to understand patterns for every record in the database, sampling should not be done.
  - If the data warehouse is very large, processing power is limited, or speed is more important than complete analysis, it is wise to draw a sample.
- If general patterns exist in the data as a whole, these patterns will be found in a sample.

Explore

- Identify relationships within the data.
- After the data are sampled, the next step is to explore them visually or numerically for trends or groups.
  - Both visual and statistical exploration (data visualization) can be used to identify trends.
  - The researcher also looks for outliers to see if the data need to be cleaned, cases need to be dropped, or a larger sample needs to be drawn.

Modify

- Based on the discoveries in the exploration phase, the data may require modification.
- Clustering, fractal-based transformation, and the application of fuzzy logic are completed during this phase as appropriate.
- A data reduction program, such as factor analysis, correspondence analysis, or clustering, may be used.
- If important constructs are discovered, new factors may be introduced to categorize the data into these groups.
- In addition, variables based on combinations of existing variables may be added, recoded, transformed, or dropped.
- At times, descriptive segmentation of the data is all that is required to answer the investigative question.
    - If a complex predictive model is needed, the researcher will move to the next step of the process.

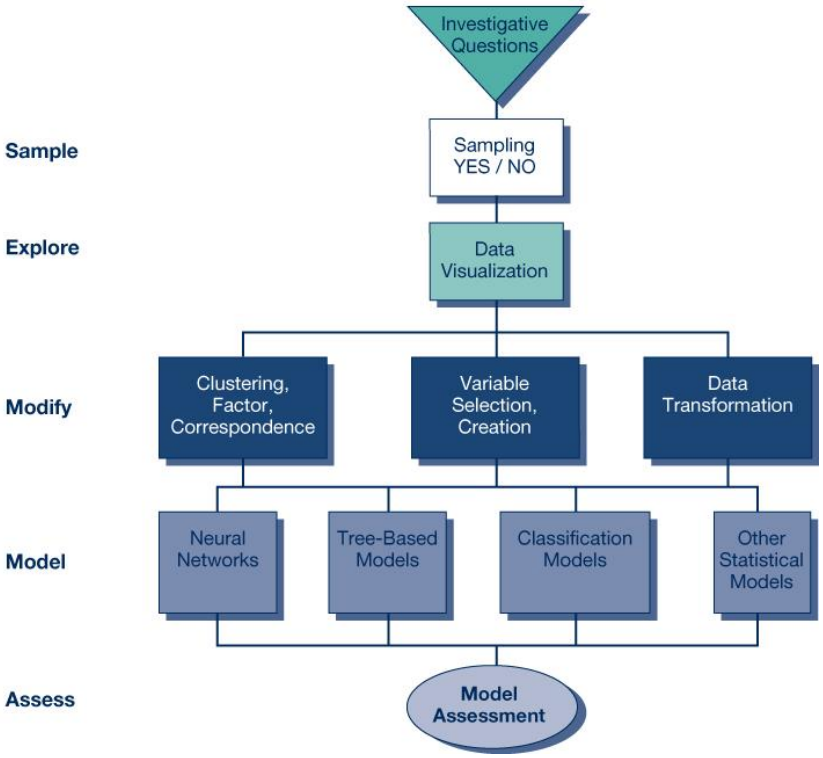
Model

- Once the data are prepared, construction of a model begins. The model should explains the data relationships.
- Modeling techniques include: neural networks, decision trees, sequence-based classification and estimation, and generic-based models.

Assess

- The final step in data mining is to assess the model to estimate how well it performs. Test the model's accuracy.
- A common method of assessment involves applying the model to a portion of data that was not used during the sampling stage.
    - If the model is valid, it will work for this "holdout" sample.
    - Another way to test a model is to run the model against known data.

Example: If you know which customers in a file have high loyalty and your model predicts loyalty, you can check to see whether the model has selected these customers accurately.



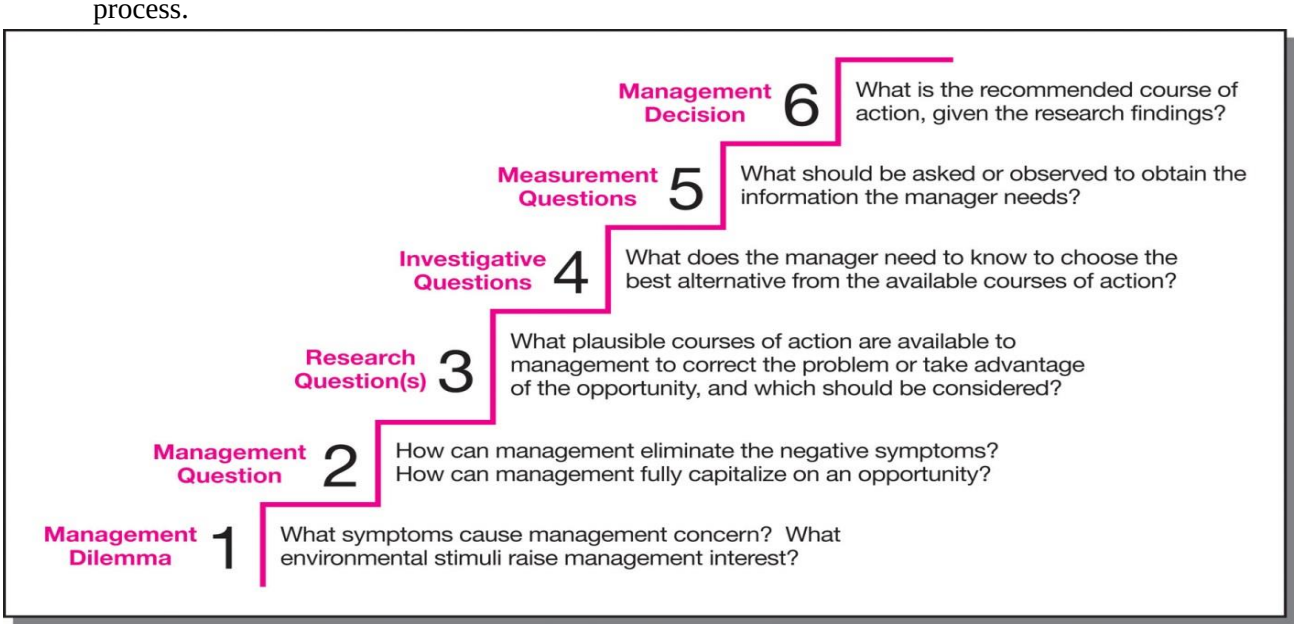
THE QUESTION HIERARCHY: HOW AMBIGUOUS QUESTIONS BECOME ACTIONABLE RESEARCH

- The process we call the management-research question hierarchy is designed to move the researcher through various levels of questions, each with a specific function within the overall business research process.

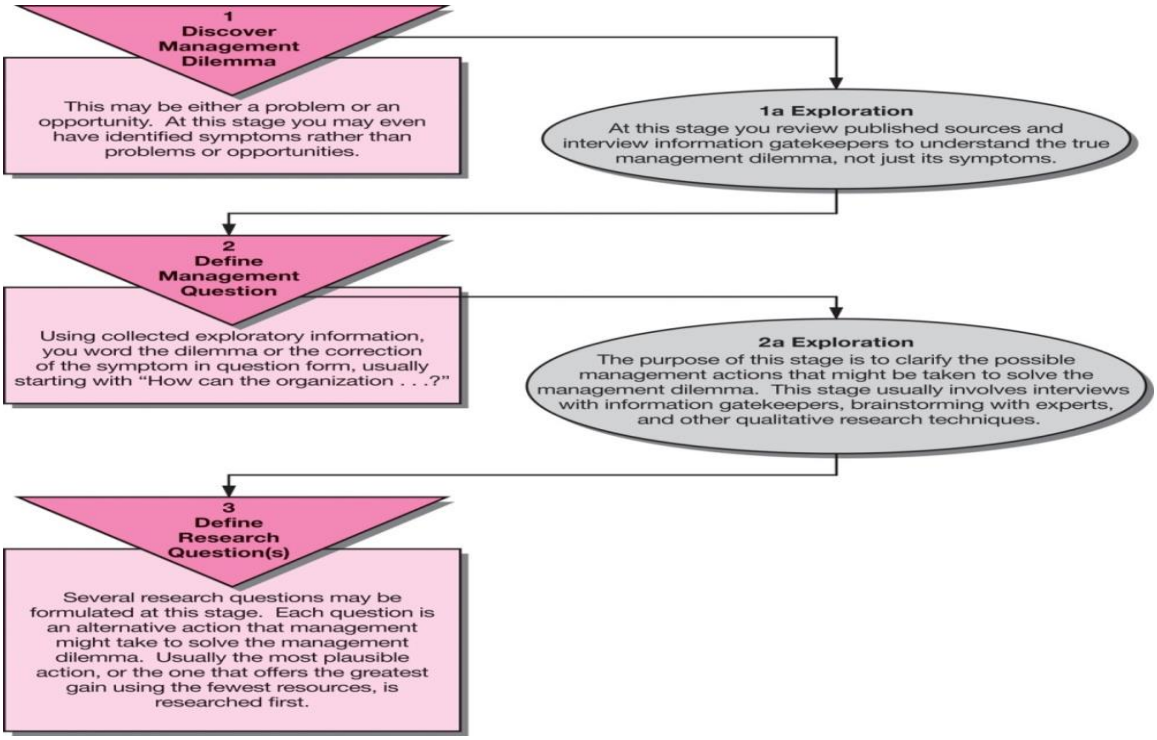
❖ The **management question** is seen as the management dilemma restated in question format.

❖ A **research question**(s) is the objective of the research study.

- It is a more specific management question that must be answered.
- Incorrectly defining the research question is the fundamental weakness in the business research process.



- Fine-Tuning the Research Question.
  - Fine-tuning the question is precisely what a skillful practitioner must do after the exploration is complete.
  - At this point the research project begins to crystallize in one of two ways:
    - It is apparent the question has been answered and the process is finished.
    - A question different from the one originally addressed has appeared.



- Other research-related activities that should be addressed at the fine tuning stage are:
  - Examine the variables to be studied.
  - Review the research questions with the intent of breaking them down into specific second-and third-level questions.
  - If hypotheses (tentative explanations) are used, be certain they meet the quality test.
  - Determine what evidence must be collected to answer the various questions and hypotheses.
  - Set the scope of the study by stating what is NOT a part of the research question.

❖ **Investigative questions** are questions the researcher must answer to satisfactorily arrive at a conclusion about the research question.

- Typical investigative question areas include:
  - Performance considerations.
  - Attitudinal issues (like perceived quality).
  - Behavioral issues.

❖ **Measurement questions** are the questions asked of participants or the observations that must be recorded.

- Measurement questions should be outlined by the completion of the project planning activities.
- Two types of measurement questions are common in business research:
  - **Predesigned measurement questions** are questions that have been formulated and tested previously by other researchers.
    - Such questions provide enhancement validity and can reduce the cost of the project.
  - **Custom-designed measurement questions** are questions formulated specifically for the project at hand.

✓ **These questions are collective insights from all the activities in the business research process completed to this point, particularly insights from exploration.**