Med Inf 405
Session 8 - Study Questions and Answers

1. **What are the components of a framework used to assess the informational value of data?**
   Data dimensions, Aligning data quality w buss. practices, Identifying authoritative sources & integration keys, Merging models, Uniting updates of varying frequency & overlapping or gapped data sets.

2. **What type of source contains the most reliable value for a specific data element?**
   Authoritative

3. **What do defined data dimensions help do with regards to the data model?**
   A frame of reference for charact. relationships between data sets & for defining metadata required for unique placement of a data set among others. Models & dimensions become critical to ensure meaning & testable results; integration of patient-dependent knowledge & external knowledge bases.

4. **What happens to data during the Extract-Transfer-Load process?**
   Data is extracted from source system (push or pull); Data is transformed to satisfy business requirements; Transformed data is then loaded into target database. Data are extracted, merged, cleaned & validated.

5. **What does the conceptual data model identify and define?**
   It assesses the informational value of data and includes ident required DEs, finding authoritative sources & creating a strategy for their integration. CDM addresses one of the key issues in intelligent data analysis: CDM defines the variables that are available for the study & those that are outside of it.

6. **What layers does integration occur at?**
   4 major layers: data sources, DEs, Data Sets, Data Values. This includes integratn of concepts, models, controlled vocabularies, methods of data acquisition, frequencies of updates, as well as units & formats of records.

7. **What is the most important characteristic of a data source?**
   Data Source

8. **What are the characteristics of a focal data element?**
   Focal DEs are usually mandatory & have the highest quality due to various validation techniques applied to them.

9. **Can peripheral data elements be integrated effectively?**
   No. They don't support any business functions, not validated, & therefore not reliable /misleading. Meaningful integration may occur only between sources with a shared pool of focal DEs.

10. **What does a controlled vocabulary provide to a project?**
    A set of unambiguously defined terms. Provides quality, & consistency in data collection, processing & interpretation within a project.

11. What categories can data elements be split into for the purposes of integration?
Integration Keys, Informative Des and Auxiliary Des.

12. What type of data elements can integration keys and informative data elements be chosen from?
Integration keys & informative data elements can be chosen only from focal data elements. Auxiliary DEs can be obtained from either focal or peripheral DEs.

13. What is an integration key?
A combination of Des that identifies exactly the same entity in two sources and is chosen from the overlapping focal Des.

14. What is the most crucial part of an integration project?
Choosing integration keys.

15. What are the most common ways of acquiring data from a source?
Data gathering agents, data pulls, & flat file updates.

16. What is challenging when performing the semantic integration of models?
Identifying an authoritative data source for every DE & multiple ways to interpret relationships between entities.

17. What does schema matching involve?
Database re-engineering, schema transformations & middleware data models. It remains a manual, labor-intensive & tedious process.

18. What type of mapping of data elements is the most frequently used method for data exchange?
Direct system-to-system mapping of data elements.

19. What are the components of the information pipeline architecture?
Multiple Data Sources→Operational Data Store (ODS) → Extract-Transfer-Load (ETL)→Core DB→Multiple Data Marts→End user functions of reporting, visualization, analysis, & data mining.

20. What is the process to identify authoritative data sources and integration keys?
Define the conceptual data model (CDM); Define the required DEs; Find sources that contain the required DEs; Identify focal DEs within these sources; Find other sources & find an authoritative source for every required DE; Find additional sources to ensure overlap in focal DEs (look for foreign keys and identify them as integration keys; Do auxiliary DEs need to be added to support data matching, if yes go find some, & choose the authoritative data source for every DE and proceed with integration.