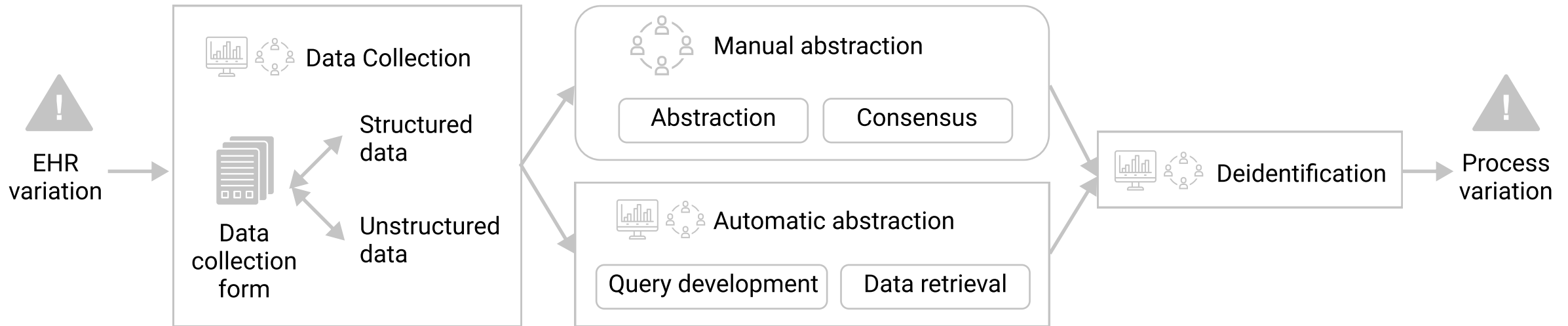# 2. Data Collection

Recommendations for Data Quality

EHR system variation

- Compare data type, document structure, and metadata
- Conduct a semi-structured interview to obtain information about the context of use
- Data quality assessment, e.g., chart review, cross source examination

Process variation

- Standardize ETL process, data linkage methods, data identifiers

Recommendations for Process Documentation

Data collection form

- Data identifiers, e.g., document id, document date, and patient clinic number
- Data definitions, e.g., how is data measured, are there any composite definitions
- Metadata of ETL: process logs and additional information about when the data is created, moved, modified, or filtered
- Abstraction method: steward, tools (e.g., REDCap), and data collection methodologies (e.g., API, SQL, standardized tools)