# Data Integrity
# Data Management

Pravich Tanyasittisuntorn, MD

# Data Quality
# Data Integrity

# Data Quality: องค์ประกอบคุณภาพข้อมูล

**ALCOA**

① **A**ccurate

② **L**egible (human readable)

③ **C**omplete & **C**ontemporaneous (recorded at the time the activity occurs)

④ **O**riginal (original sources or primary data)

⑤ **A**ttributable to the person who generated the Data

① ถูกต้อง

② อ่านออก

③ ครบสมบูรณ์ และ
บันทึกข้อมูลทันทีที่ได้รับ

④ ข้อมูลจากแหล่งกำเนิดโดยตรง

⑤ สามารถระบุเจ้าของ ผู้บันทึก
วันที่บันทึก

# Data Integrity

- **Data integrity:** process of colleting data ขึ้นกับขบวนการเก็บรวบรวมข้อมูล

  ① **Credible**: ความน่าเชื่อถือ

  supported by known facts, worthy of belief and confidence

  ② **Internally consistent**: ความสอดคล้อง และเที่ยงตรงของข้อมูลและคำที่ใช้

  non-contradictory and consistent terms

  ③ **Verifiable**: สามารถตรวจสอบความถูกต้อง

# Ethical Aspects

# Ethical Consideration

- Respect and protect confidentiality and privacy of subject participating in the research
- Questions
  A. What personal data/information is identifiable?
  B. What data is sensitive or non-sensitive?
  C. What is the impact if the personal data is disclosed or the subject's confidentiality is breached?

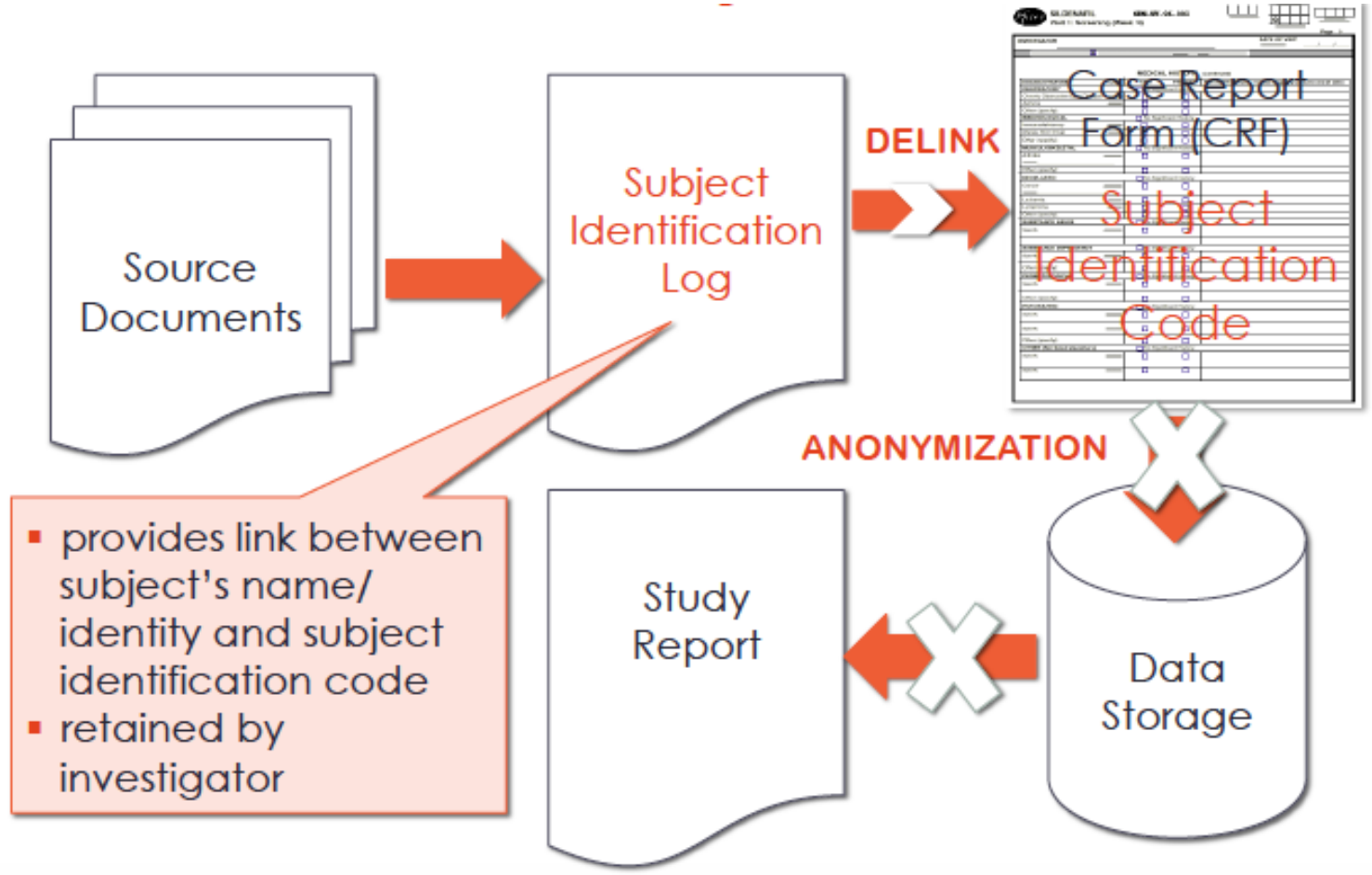
# Ethical Consideration

- How to protect research subjects' confidentiality and privacy?
  1. Protocol
     - Specify data/variables to be collected,
     - Data collecting process: how data is collected, analyzed and reported
  2. Collect minimal identifiable data/information
  3. Protocol obtained IRB/IEC approval
  4. Conduct and collect the data in compliance with the approved protocol

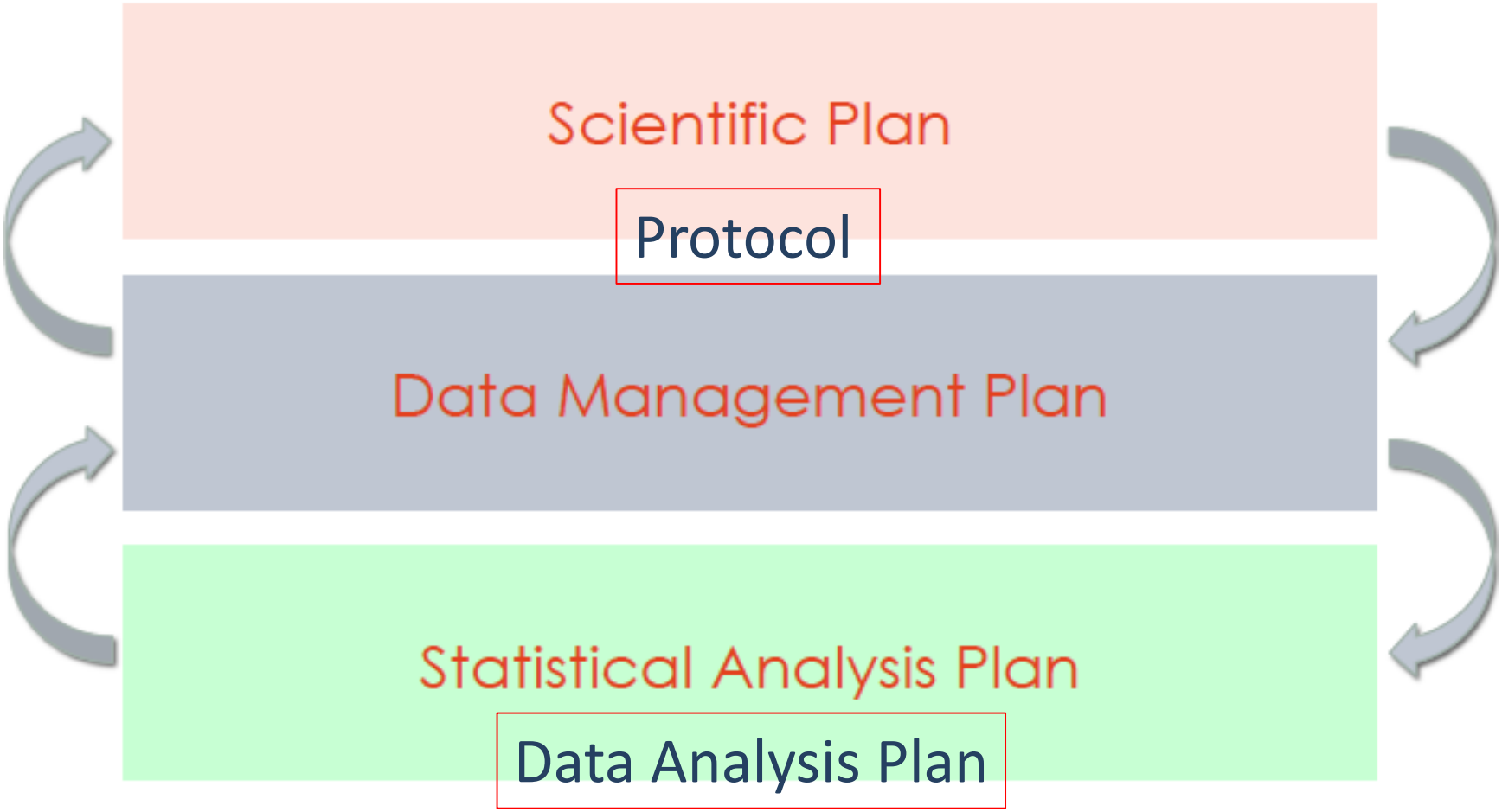
# Ethical Consideration

- How to protect research subjects' confidentiality and privacy?
  - ① Establish confidentiality protection in process of data collection, data management, data analysis and reporting study results (delink and anonymized process)
  - ② Inform consent process
    - subjects the data collection and confidentiality protection process and ask for permission to access (in the informed consent process)

# Confidentiality Protection



# Data Collection and Data Management


Scientific Plan
Protocol
Data Management Plan
Statistical Analysis Plan
Data Analysis Plan

# Data Flow and Data Quality Control

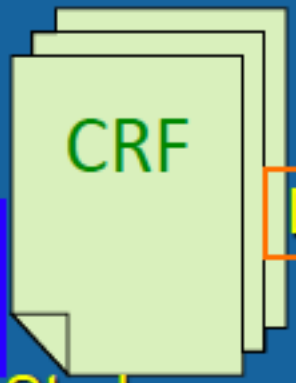

Investigator

Data clarification

Data Clarification Form

Self-evidence correction

Source Data Verification




Data validation

Data entry

Source document

Study nurse/
Site coordinator

Investigator review & sign

Data management

# Data Management Plan

1. Data collection tool:
   - CRF (case report form)
   - Questionnaire for survey study
2. Source document plan:
   - Source documents and flow
   - Develop worksheet, if required
3. Data storage:
   - Database structure and database application

# Data Management Plan

4. Data coding
   - Data dictionary
   - Annotated CRF
   - Textual data e.g. disease, drug, adverse event terms
5. Data flow and CRF tracking
6. Data entry:
   - Manual
   - Scanning
   - EDC (electronic data capture)

# Data Management Plan

7. Data cleaning / data validation / edit checks
   - identify
     - missing page
     - obvious data error, data inconsistency, non-meaningful data, omitted / missing data, additional hand written data
     - protocol deviation
   - time-of-entry and back-end edit checks
   - review data in CRF point-by-point and page-by-page

# Data Management Plan

7. Data cleaning / data validation / edit checks
   - Activities
     - review CRF pages
     - manually review data
     - identify discrepancies through computer generated checks / validation
     - conduct logical and statistical review checks

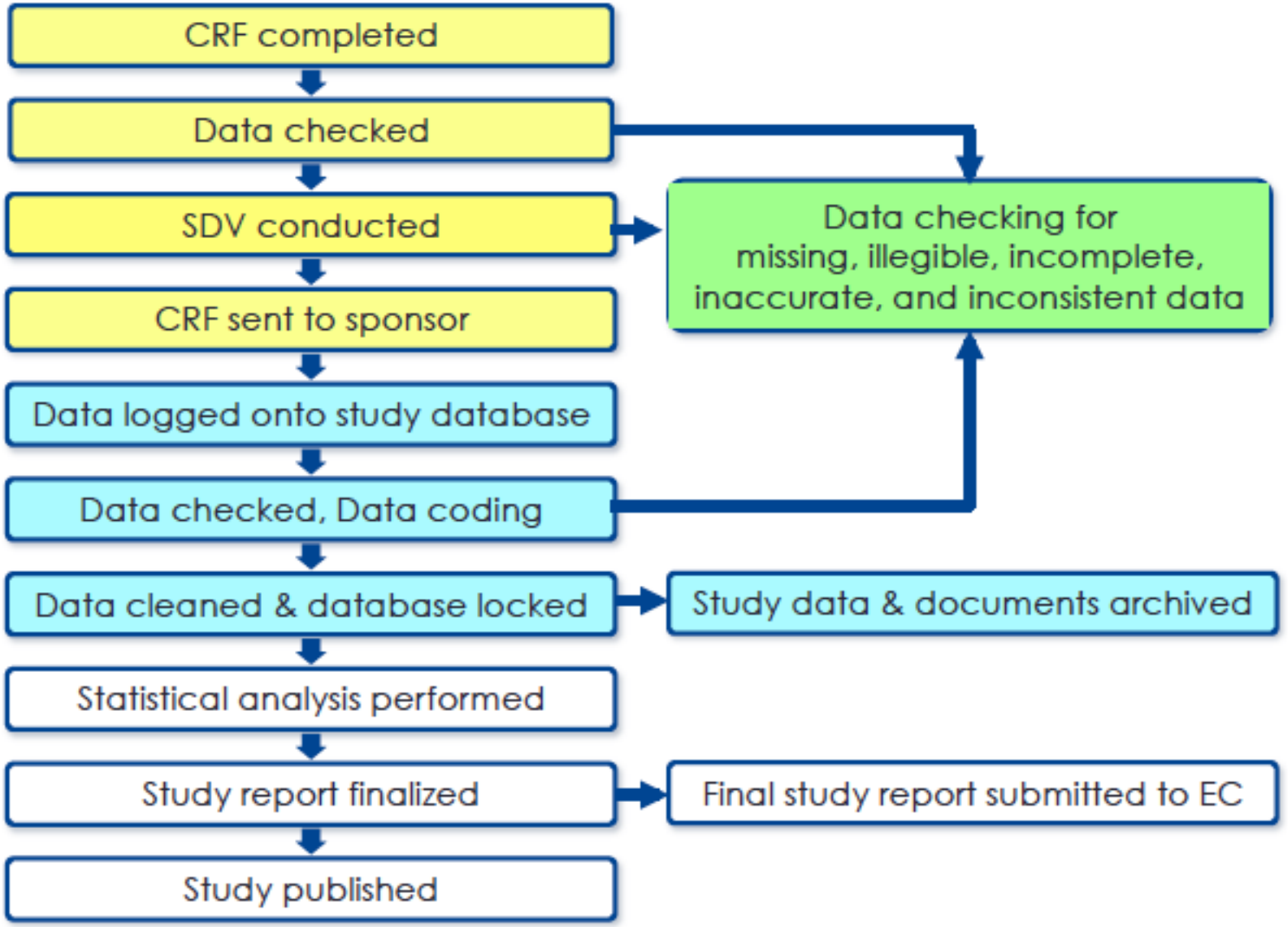
# Data Management Plan

8. Managing discrepancies
   - Query generation and resolution
   - Protocol deviation
   - Data integration
9. Managing lab data and safety data
   - Lab test unit
   - Lab normal range
   - Medical review for abnormal test
   - Data transfer, if lab database is separated

# Data Management Plan

10. Managing safety data
    - AE terms
    - Medical review for safety data completion and causality assessment
    - SAE reconciliation
    - Continual review
11. Database locked (at interim analysis and study completion)
12. Data analysis and data quality review

# Data Flow and Data Quality Control



# Statistical Analysis Plan (SAP)

a. Study objectives and variables
b. Hypothesis to be tested
c. Primary end points & secondary end points
d. Definition of analysis populations:
   - intention-to-treat, per-protocol, and safety population
e. Statistical methods to be used
   - be specific: what covariates, compare what treatment groups, in what FU visits)
f. Subgroup analysis
g. Data handling rules: missing data
h. Data presentation:
   - Tables, figures and line listings (TFLs)

# Data Handling in Qualitative Research

- Audit trail
  a) Raw data
  b) Data reduction and analysis products
  c) Data reconstruction and synthesis products - including structure of categories (themes, definitions, and relationships), findings and conclusions and a final report
  d) Process notes
  e) Materials relating to intentions and disposition
  f) Instrument development information - including pilot forms, preliminary schedules, observation formats