ICAO UNITING AVIATION

Module 2
Data

M. Merens

IAA, Air Navigation Bureau

Digitization

Event

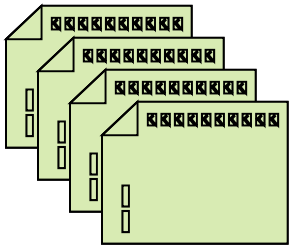
Airport

Airline

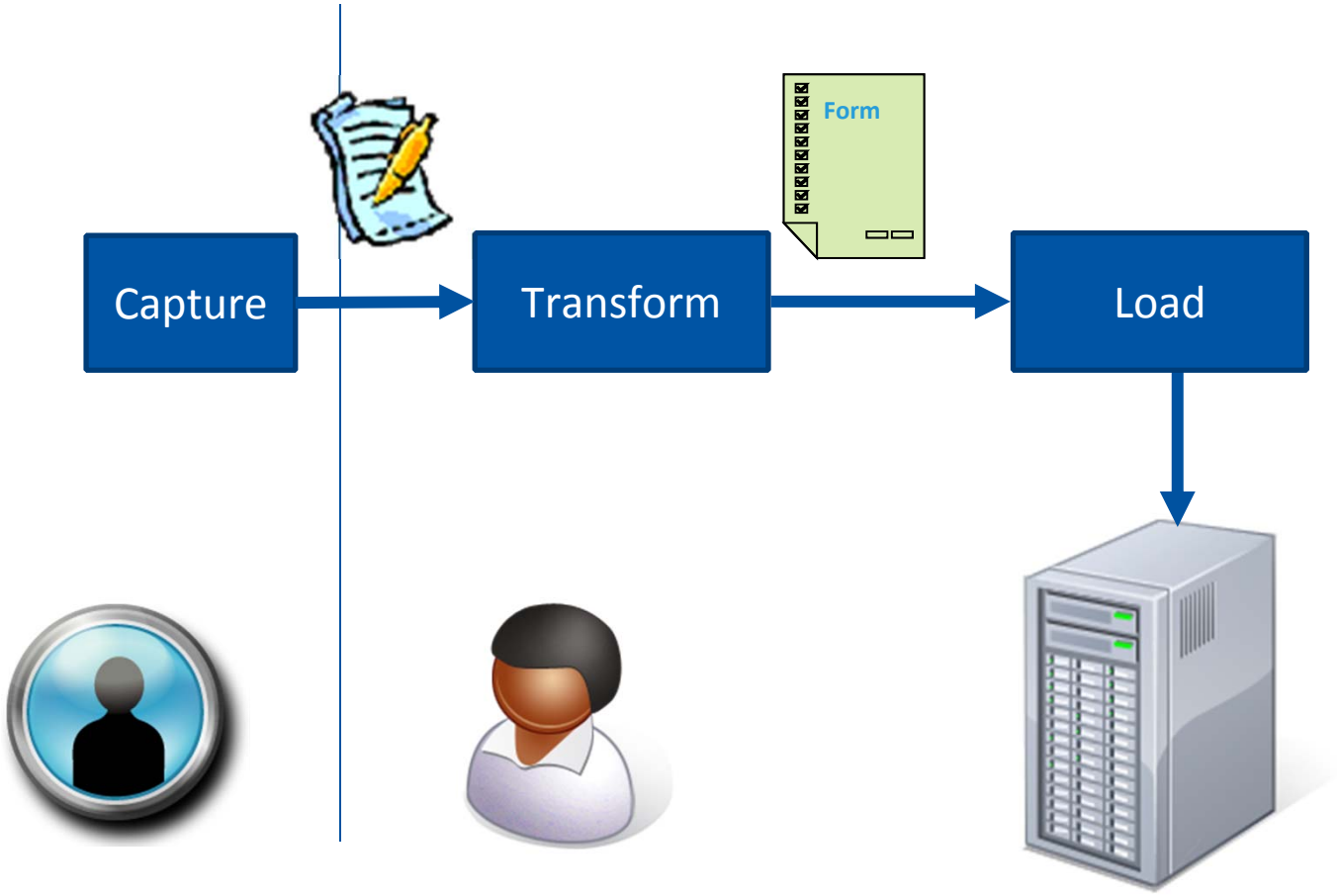
Collect characteristics about entities

Database

Entity records



Collection Process



Collection Methods

- **Proactive**
  - Surveys
  - Audits
  - Inspections
- **Reactive**
  - Reports
- **Automated**
  - Data feeds

Audits and Inspections

- An audit or inspection needs a checklist with questions
- Each question must be either Yes/Satisfactory/Good, No/Unsatisfactory/Bad or Not Applicable/Not asked
- There can only be one finding per question
- Each audit or inspections can be calculated a score as a percentage [0%-100%]

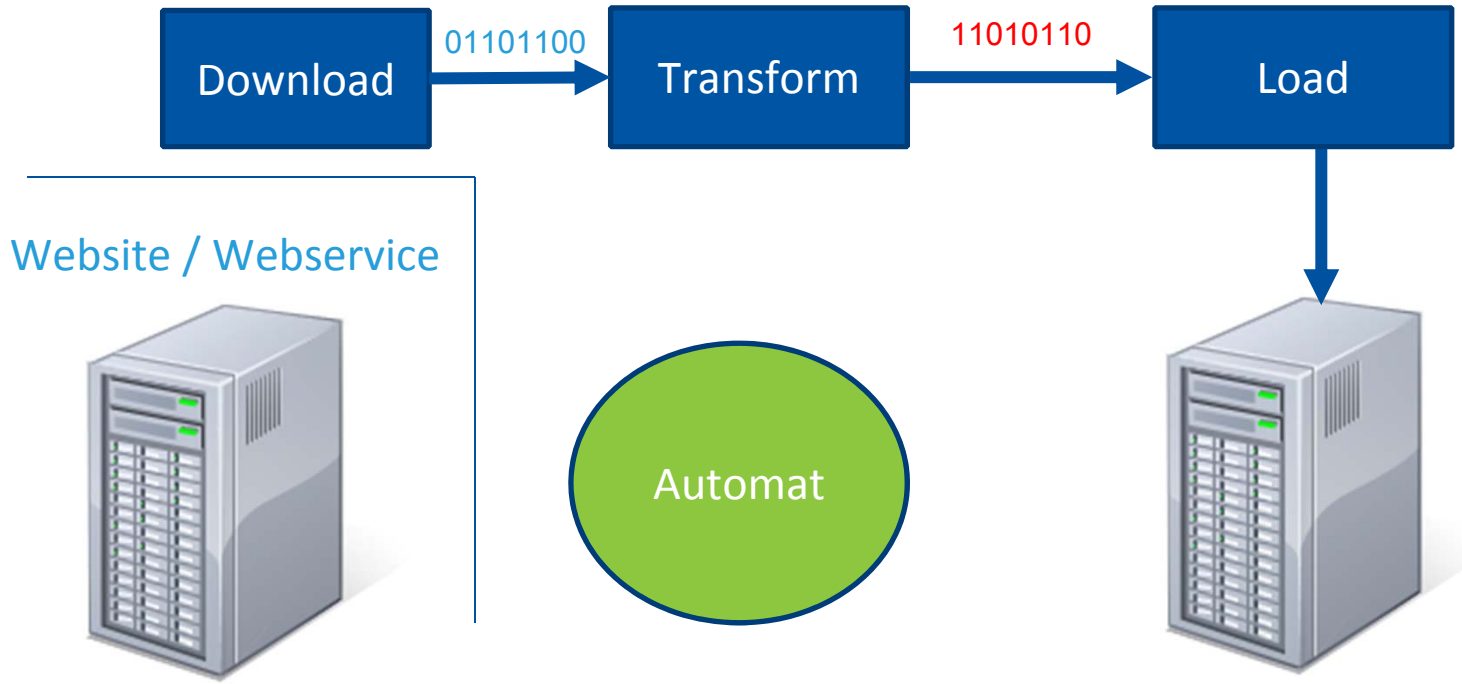
$$\text{Score} = \frac{\sum Yes}{\sum Yes + \sum No}$$

Examples

- **Good examples:**
  - USOAP Protocol Questions
  - SSP Gap Analysis Questions
- **Bad examples:**
  - EASA Part 145 audit
  - EU SAFA Inspections

Data feeds

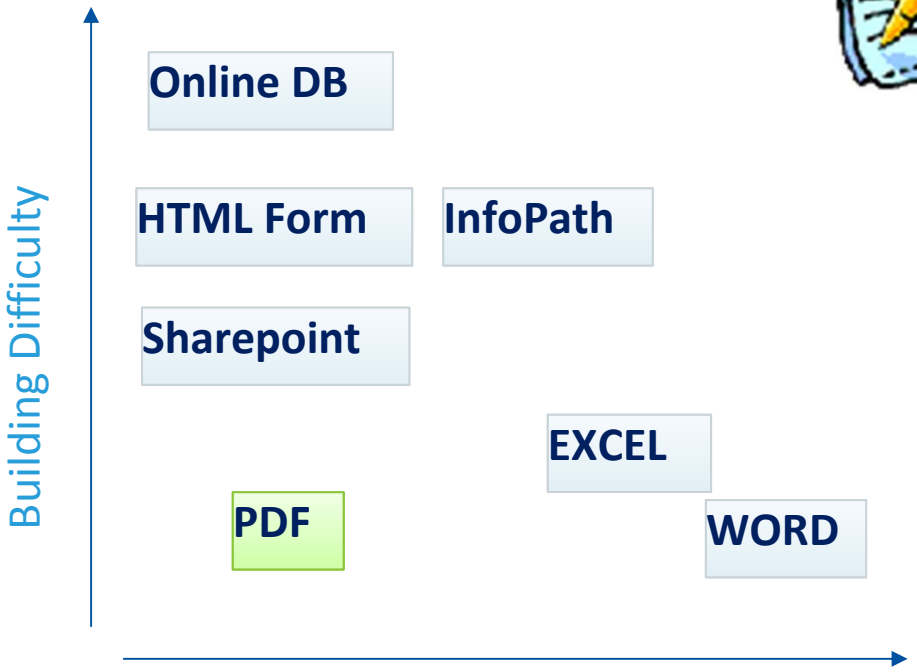
Datafeeds allow full automation of the collection process



Datafeed Sources

- **Accidents and incidents**
  - Aviation Safety Network
  - Aviation Herald
  - ASCEND
  - iSTARS ADREP
- **Traffic**
  - FlightStats
  - iSTARS State Traffic
- **ADSB Positions**
  - FlightStats
- **Flight Schedules**
  - FlightStats
- **Airport Information**
  - iSTARS Airports
  - iSTARS PBN
- **Fleet Information**
  - iSTARS Fleet
  - ASCEND / FlightGlobal
- **Terrain Data**
  - NOAA GLOBE
- **METARS**
  - NOAA

Electronic Forms



ADREP Core Taxonomy (29)

- **Filing Information (3)**
- **When (2)**
- **Where (5)**
- **Classification (2)**
- **Severity (3)**
- **Narrative (1)**
- **Aircraft identification (6)**
- **Operator (3)**
- **Operation type (1)**
- **History of flight (3)**

References

Taxonomies

ISO 3166-1

DOC7910

Occ. Class

Occ. Cat.

Acft. Damage

Injury Level

ISO 3166-1

Mass Group

ISO 3166-1

DOC8585

Op. Type

DOC7910

Flight Phase

References

Airport Codes

4 letter ICAO Code (DOC7910)

State Codes

UN 3-letter Country Codes (ISO 3166-1)

Airline Codes

3 letter ICAO Code (DOC8585)

Filing Information and Narrative

- **Basic information about who submitted the report and description**
  - Reporting Organization
  - File Number
  - Date received
  - Narrative

When and Where

- **Information about when and where the event happened**
  - UTC Date and Time
  - State of Occurrence ISO 3166-1
  - Location of occurrence
  - FIR
  - Latitude
  - Longitude

Classification and Severity

- **Classifications of the event**
  - Occurrence class

    Accident, Serious Incident,
Incident, Occurrence without safety effect, Not Determined
  - Occurrence Category

    CICTT Taxonomy
  - Damage Aircraft

    Destroyed, Substantial, Minor, None, Unknown
  - Injury Level

    Fatal, Serious, Minor, None, Unknown
  - Fatalities

Aircraft Information

- **Information about the aircraft**
  - Registration
  - Serial Number
  - Aircraft Category
    - Airplane, Helicopter, Other
  - Manufacturer / model
  - State of registry
  - Mass group
    - » MG1: 0-2250 kg
    - » MG2: 2251 - 5700 kg
    - » MG3: 5701 - 27000 kg
    - » MG4: 27001 - 272000 kg
    - » MG5: >272000 kg
    - » UNK: Unkown

ISO 3166-1

Operation

- **Information about the operator and its operation**
  - Operator Name
  - Operator Code DOC8585
  - State of Operator ISO 3166-1
  - Operation Type
    - Scheduled Commercial Air Transport
    - Non-scheduled commercial Air Transport
    - Non commercial operation

History of Flight

- **Information about the flight**
  - Last departure point DOC7910
  - Planned destination DOC7910
  - Flight Phase
    - CICTT FlightPhase taxonomy

Form Design Principles

- Only ask for data you really need to generate information
- Avoid multiple value fields (like occurrence cat.)
- Allow multiple ways to submit (print, email, online, data)
- Support multiple IT systems (MAC, MS, iPad)
- Do not ask for subjective information
- Keep the value lists limited (less than 8)

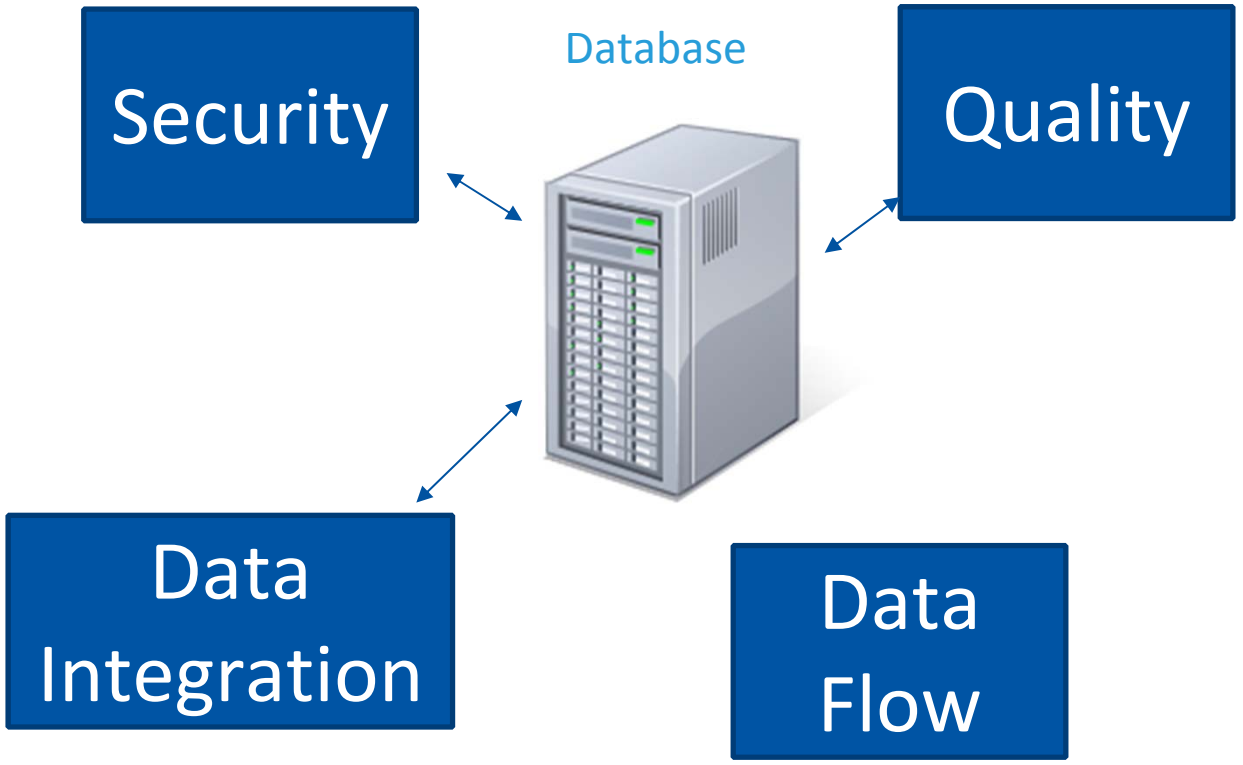
Exercise

- **Create a small PDF reporting form**
  - Go to http://www.pdfamigo.com/
  - Create a fillable PDF form for Narrative, Occurrence Class and File Information only
  - Save the form and send it to a colleague to fill it in

Exercise

- **Review existing forms**
  - Check the forms available under Report Forms
  - Which of those forms do comply to the principles and collect enough data?
  - Which do not?

Data Management



Security

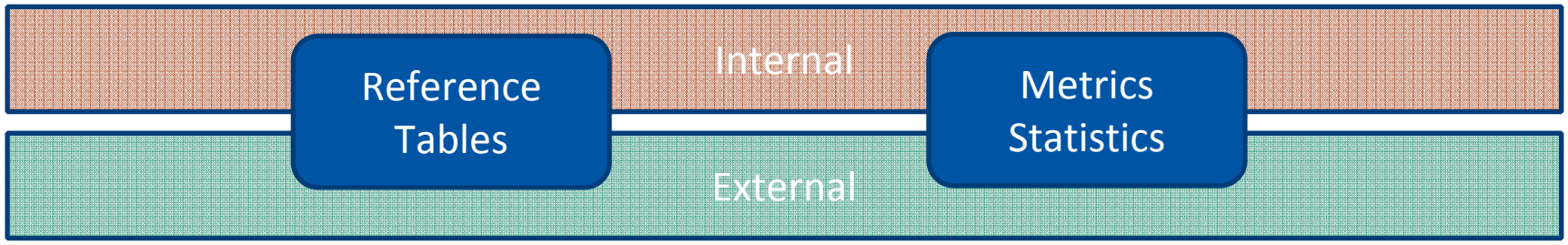
- **Threat**
  - Unauthorized people access the data
- **Mitigation**
  - Control access to the data (username/password)
  - Do not send data by email but provide downloads (secure website, dropbox, iDrive etc.)

Data Quality

- **Correctness**
  - Is the data provided correct?
- **Completeness**
  - Are all fields provided, not empty?
- **Coherence**
  - Are fields standard between records?
- **Uniqueness**
  - Are there copies made of the data?

Single Source of Truth

Single Source Of Truth (SSOT) refers to the practice of **structuring information** models and associated schemata such that **every data element is stored exactly once** (e.g., in no more than a single row of a single table).



Data Connections

Somefile.csv
original

Changes do not propagate

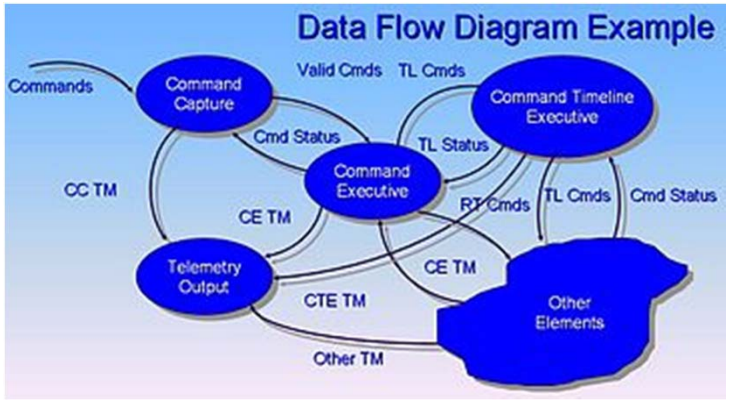
Somefile.csv
Download copy

Changes propagate

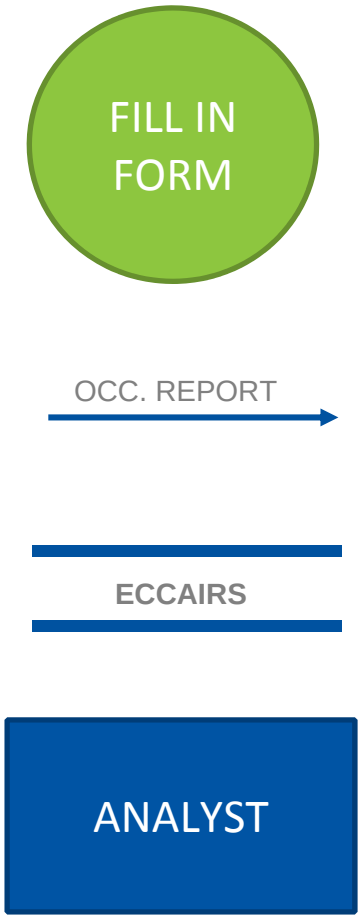
Somefile.csv
Data connection

Data Flow Diagram

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system



E. Yourdon DFD Methodology



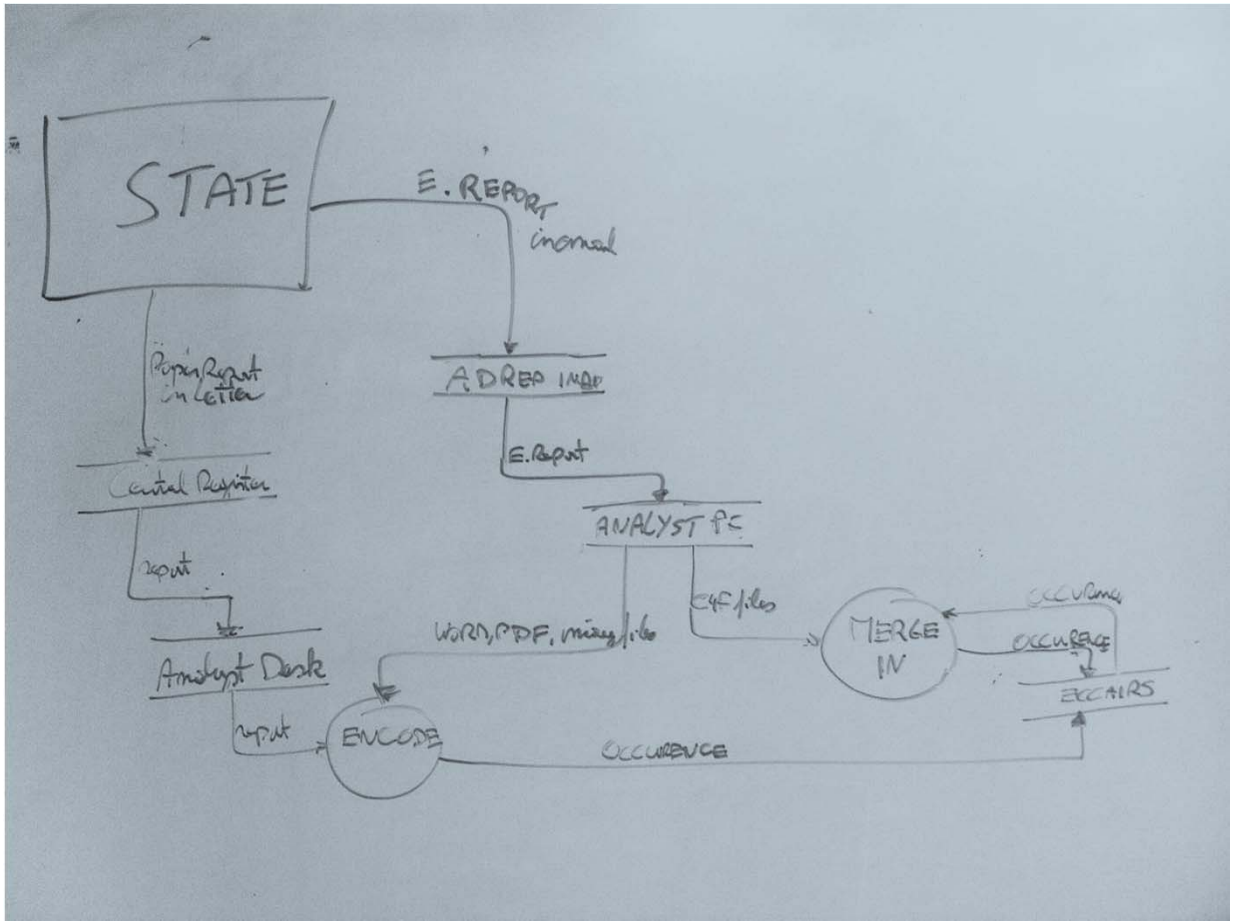
Process
The process shows a part of the system that transforms inputs into outputs

Flow
The flow is used to describe the movement of chunks from one part of the system to another part

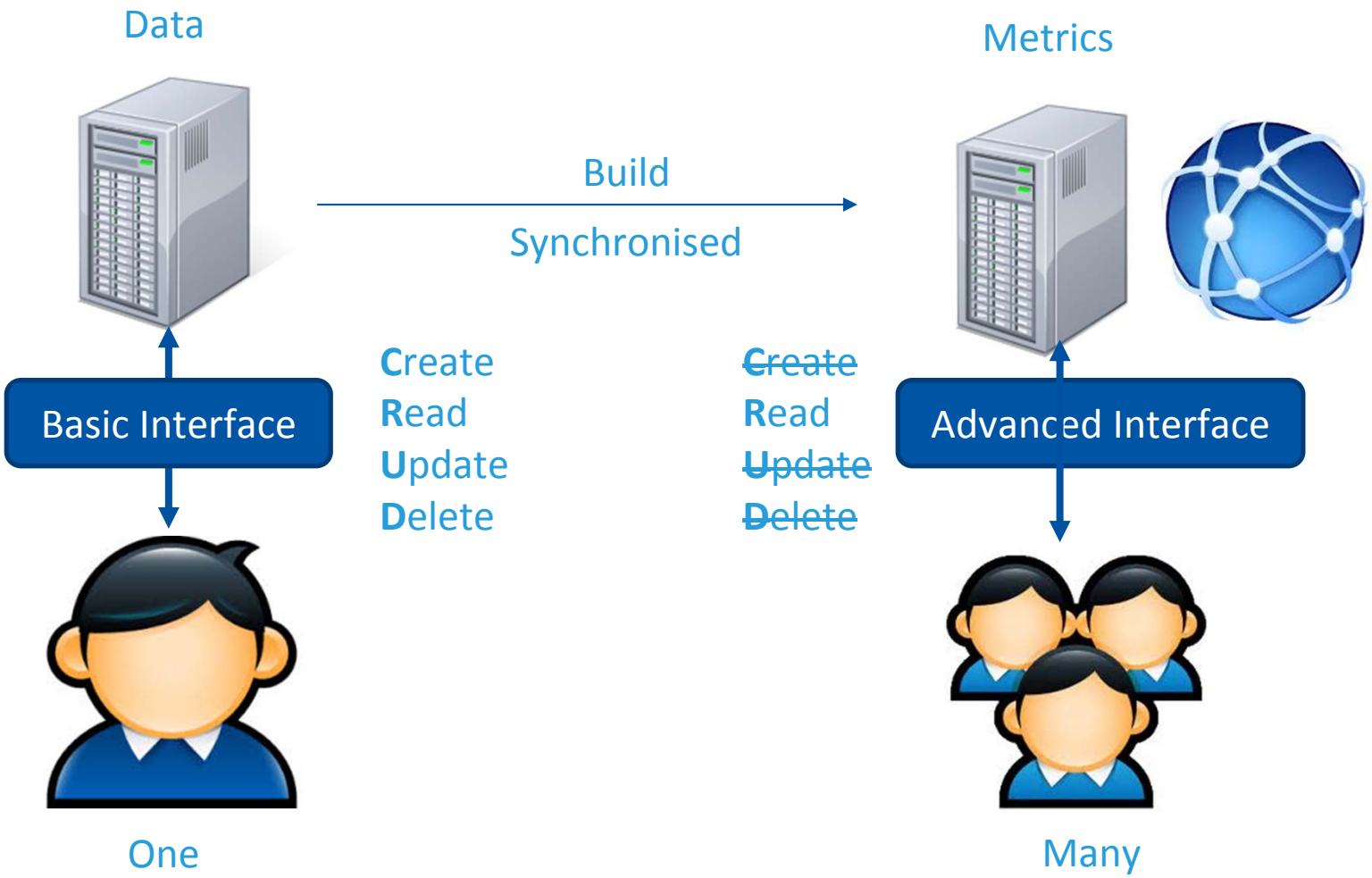
Store
The store is used to model a collection of data packets at rest like a database

Terminator
Terminators represent external entities with which the system communicates

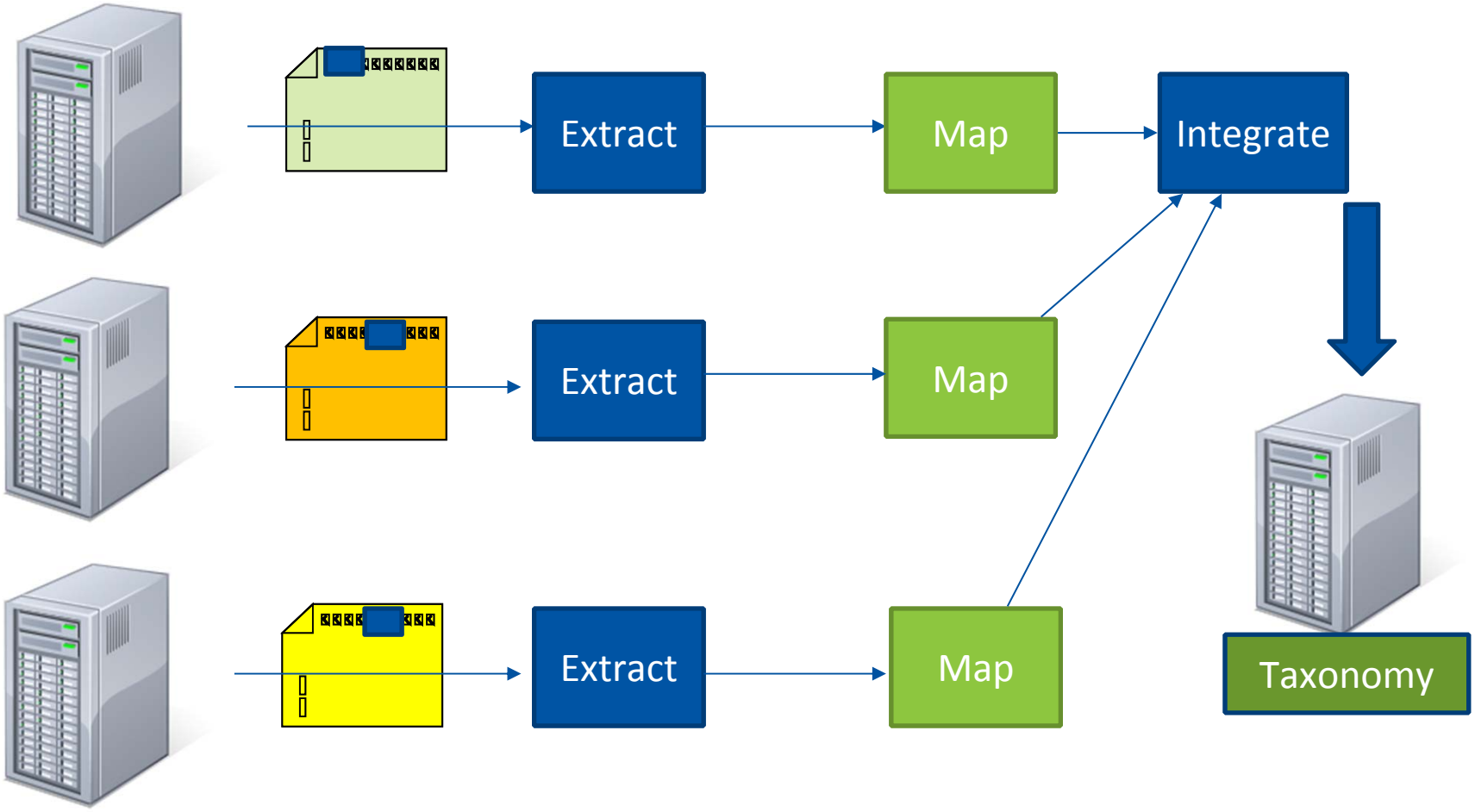
ICAO Internal Accident Reporting DFD



Metrics and Data

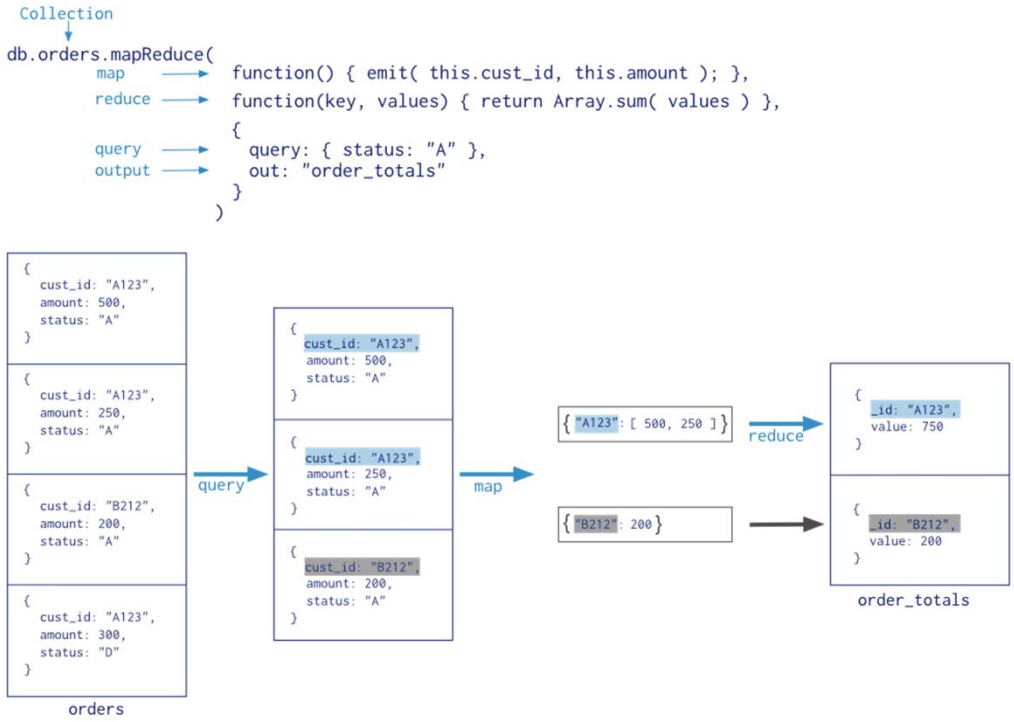


Data Integration

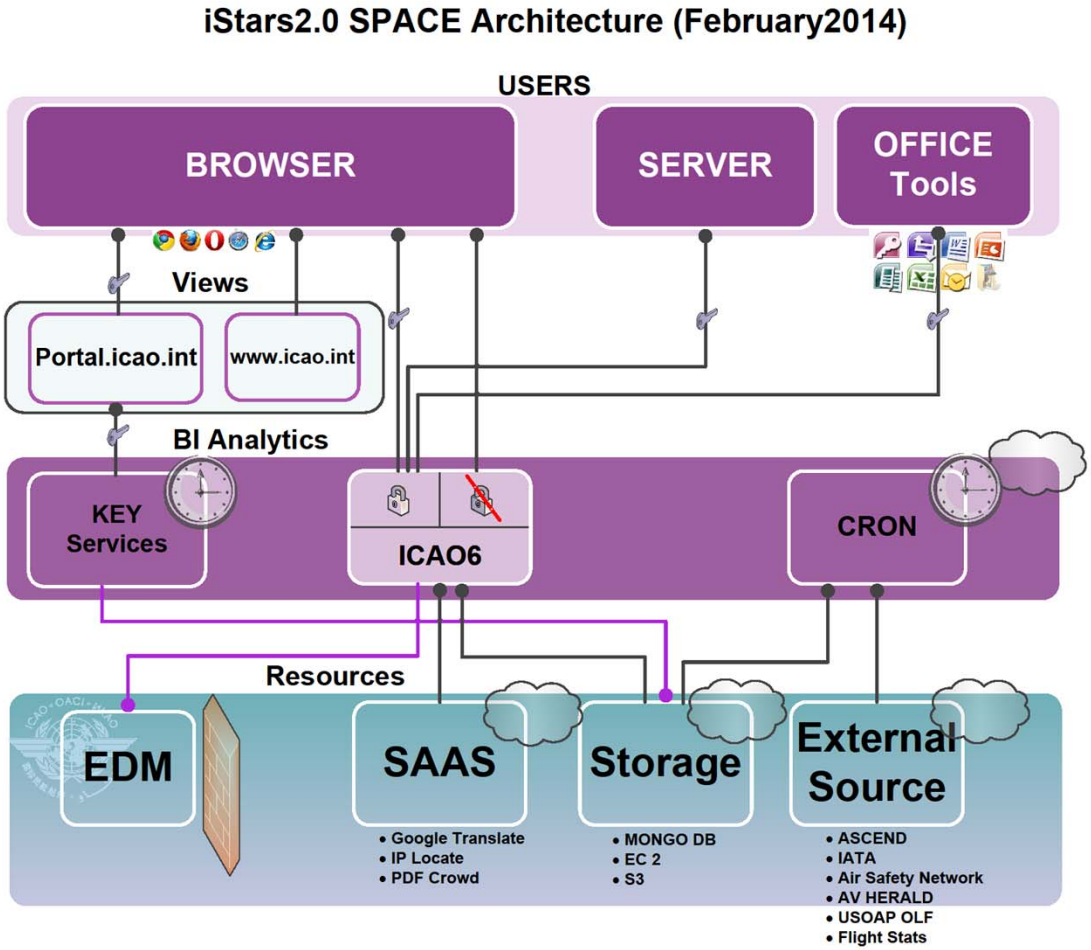


Map Reduce

- MapReduce helps building metrics out of data directly
- Can deal with any amount of data



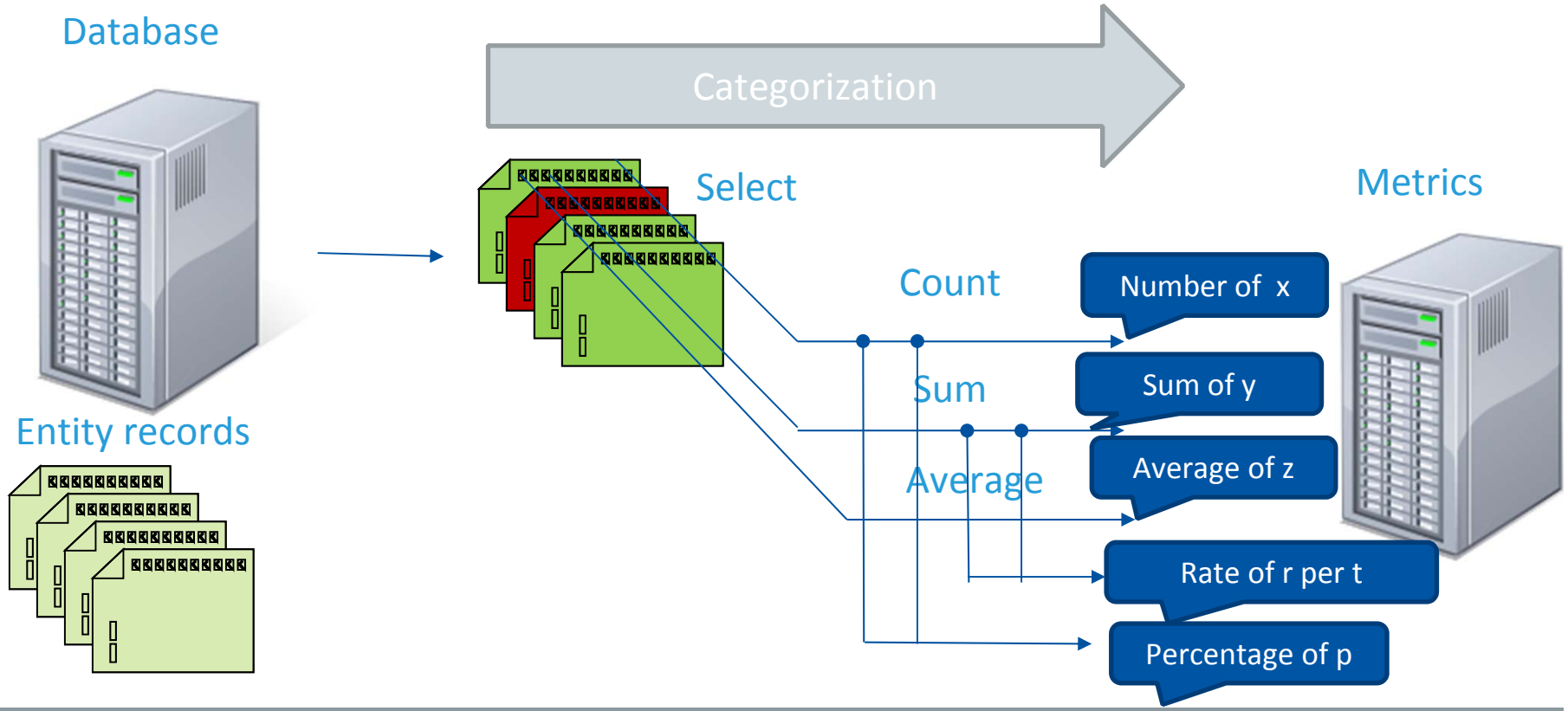
iSTARS Architecture



Exercise

- **Describe the data integration flow**
  - Draw a DFD of the process on how in your State, you complete accident data, get traffic data and calculate accident rates
  - Identify the manual interventions and possible corruptions

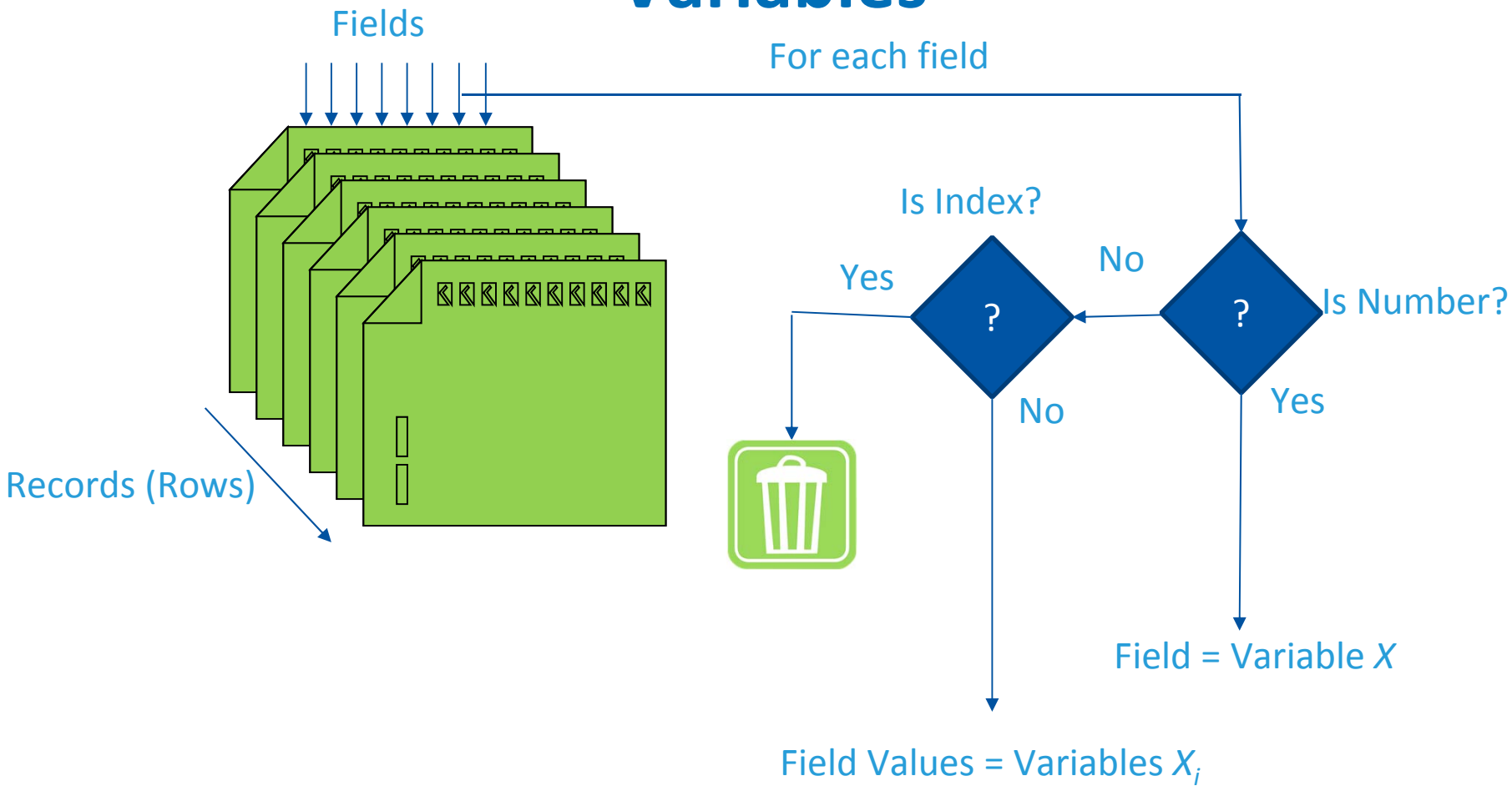
Building Metrics



Metric Types

- **Metrics are Values of Variables**
  - Percentage of
  - Count / Number of
  - Sum of
  - Average of
  - Rank of
  - Maximum of
  - Etc.

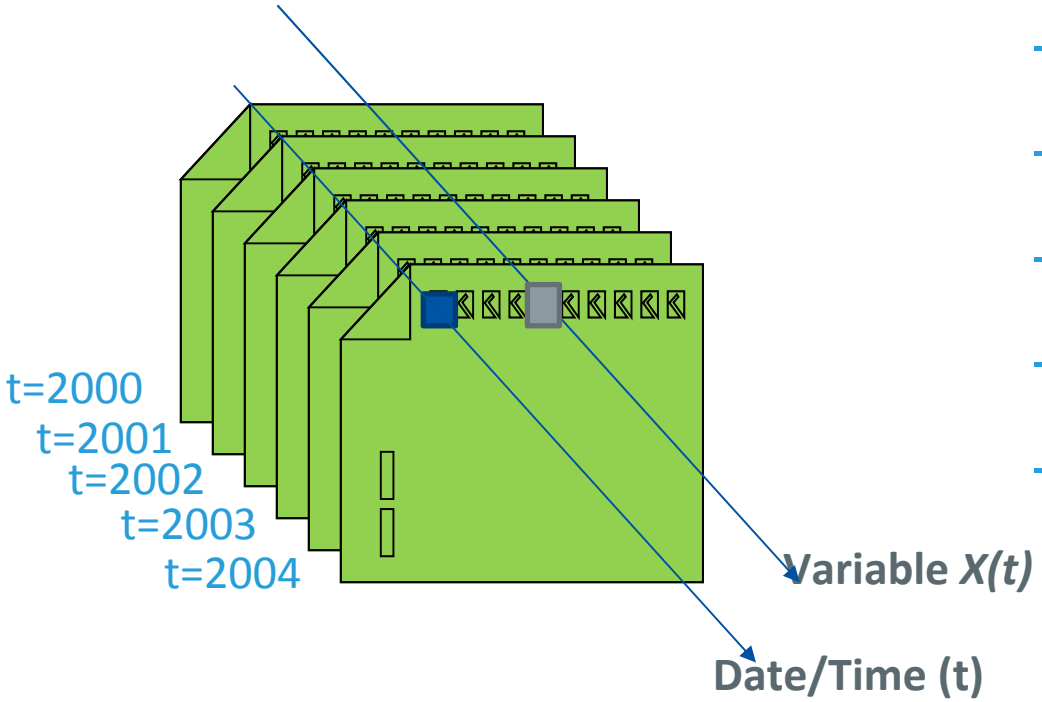
Variables


Fields
For each field
Is Index?
Yes
?
No
?
Is Number?
No
Yes
Records (Rows)
Field = Variable X
Field Values = Variables X_i

Time Series

Signal theory metrics

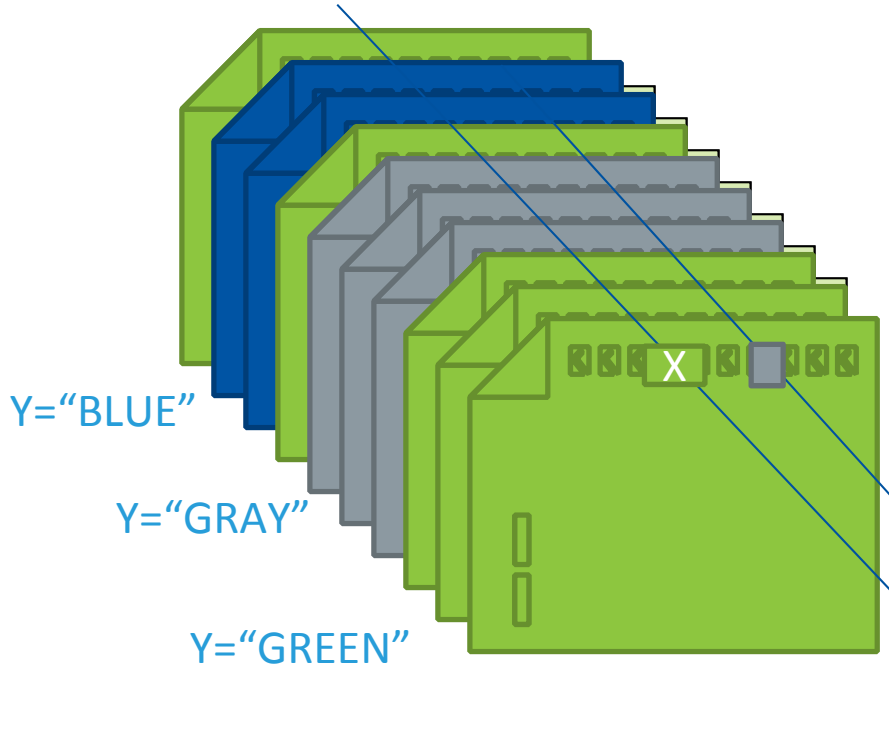
- Sliding Average
- Trend/Growth of
- Frequency of
- Bandwith of
- Signal to noise ratio of



Multi-variate fields

- **A multi-variate field is a field with more than one value:**
  - Occurrence categories: [SCF-PP;ARC;RE;F-POST]
- **For simple analysis, create one variable X_i per field value which are not globally coherent:**
  - Sum of X_i != Number of Records
- **For advanced analysis, map combinations to new single value field**
  - [SCF-PP;ARC;RE;F-POST] = “Runway Safety”

GROUP BY

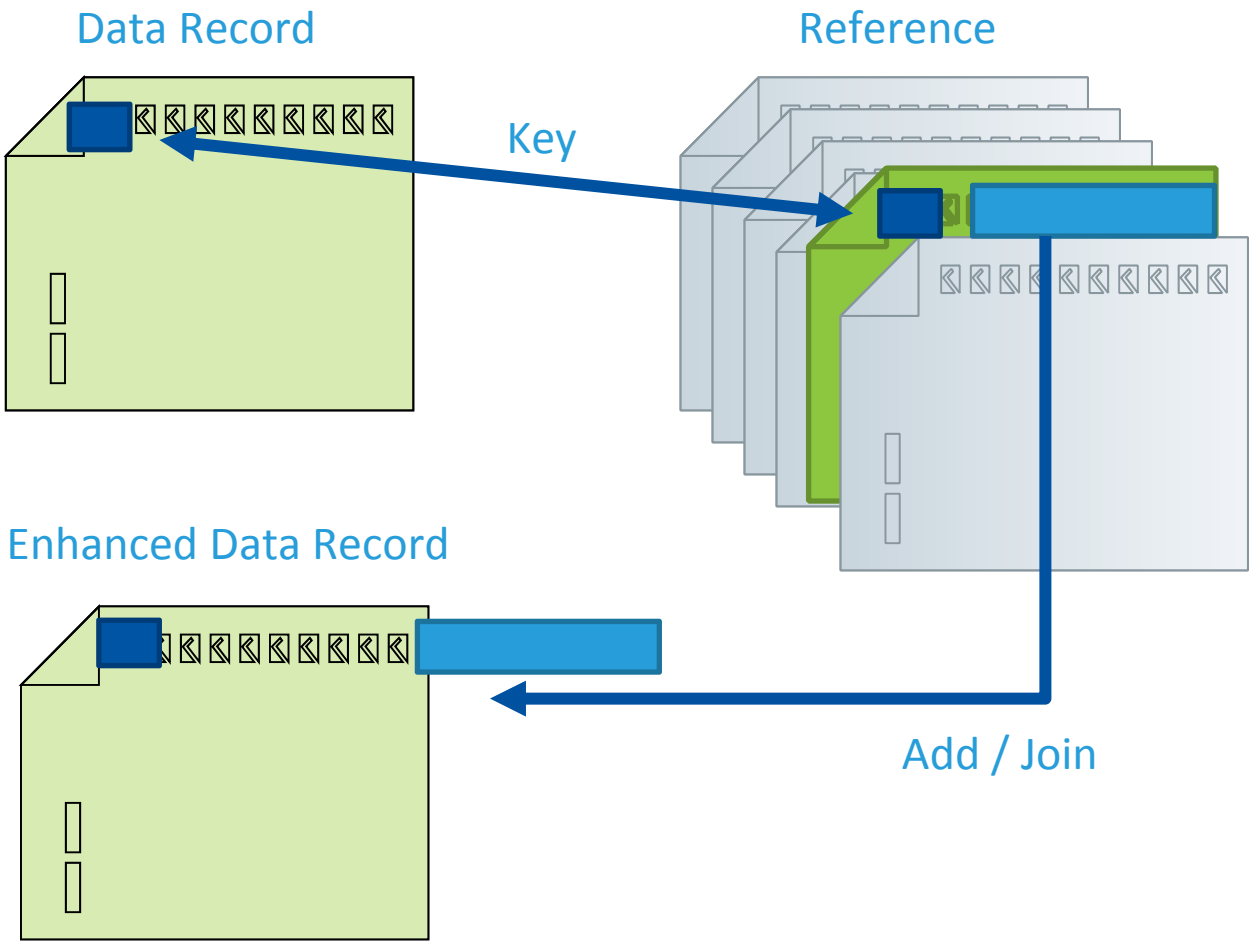


- **Apply a metric to a group of records**
  - Number of Accidents (X) by Country (Y)

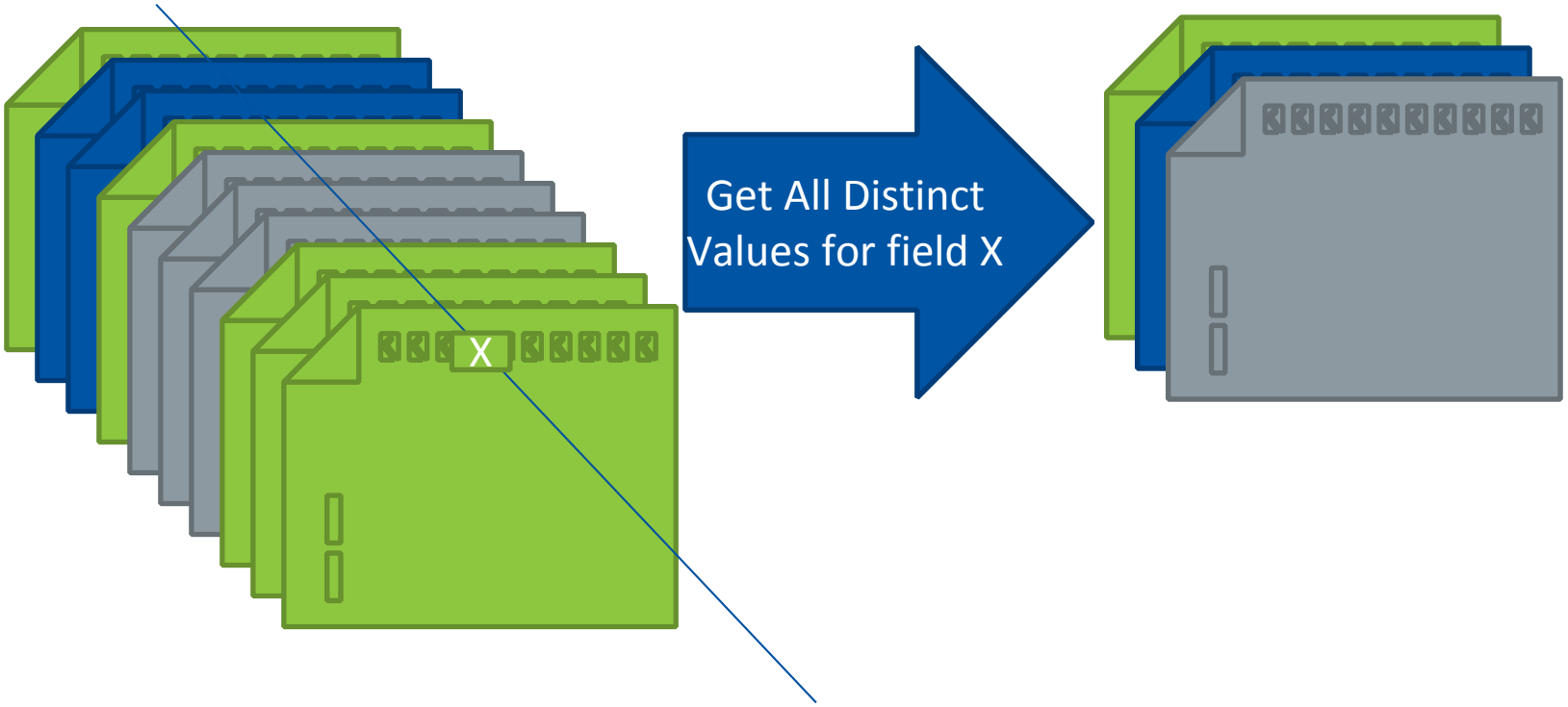
Empty Value Check

- **Empty fields represent unreported data, which is different from "Unknown"**
- **For each field, check for the proportion of records with empty values**
- **If the proportion of empty values is too high:**
  - Discard the metric,
  - Fill in the missing values, or
  - Query a subset of data to exclude the empty values

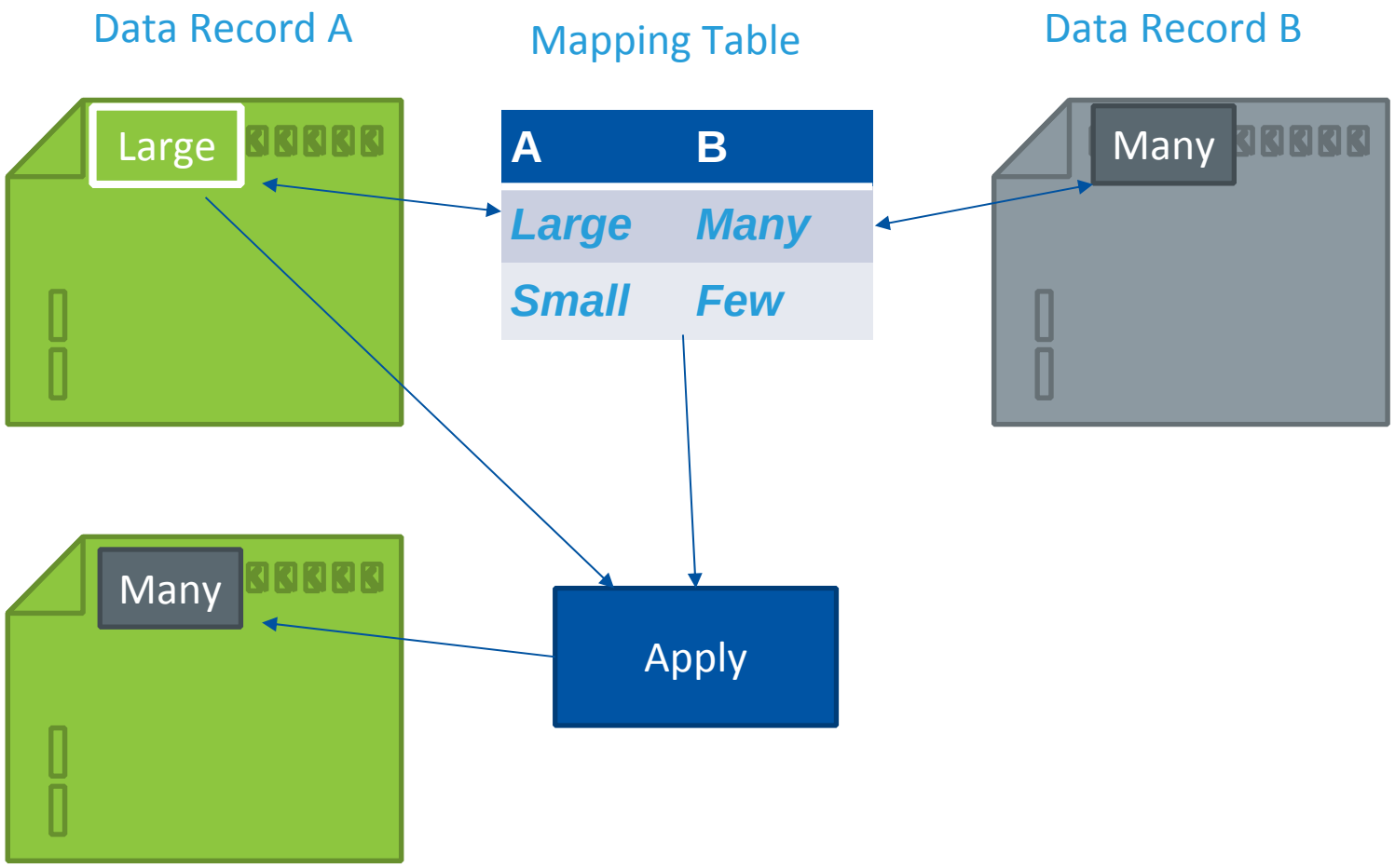
EXCEL VLOOKUP



EXCEL DISTINCT Values



Mapping Tables



Exercise

- **Practice EXCEL data query and filter**
  - Join information about States from different tables:
    - Open the Statenames file
    - Add the LEI Map World and the PBN Implementation World
    - Add to each ICAO Member State, the Overall EI and the % of PBN runways
    - What is the coverage of the data?

ICAO

THANK YOU