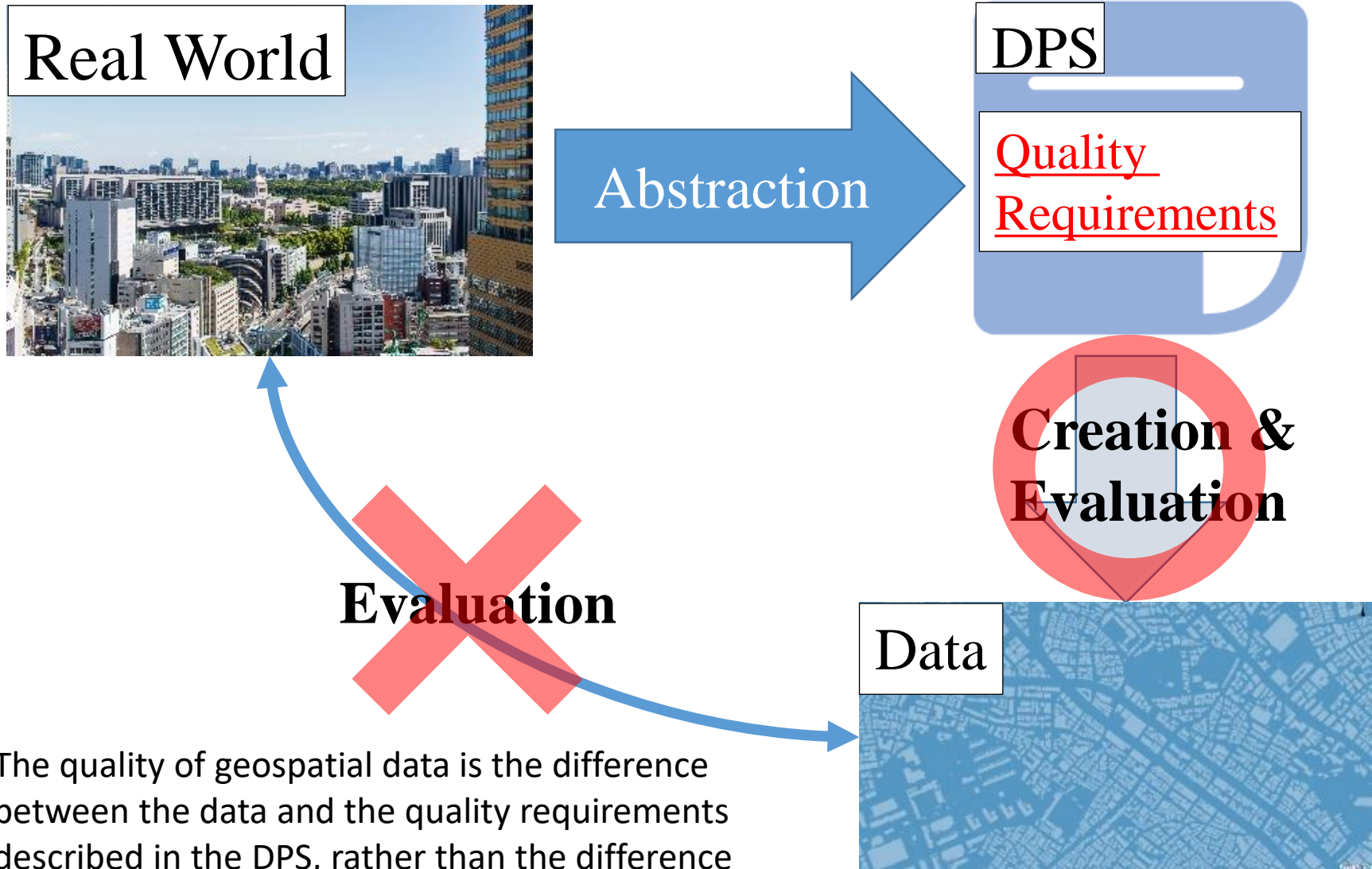


# **Basic Concept of Data Quality – requirements, evaluation and reporting –**

HATORI Tomohiko  
(JICA Expert)  
10 August 2020

# 1. Data Quality for Geographic Information



The quality of geospatial data is the difference between the data and the quality requirements described in the DPS, rather than the difference between the data and the real world.

## 2. Data Quality Elements (Quantitative)

| #  | Data Quality Elements                  | Categories          |
|----|--|---------------------|
| 1  | commission                             | Completeness        |
| 2  | omission                               |                     |
| 3  | conceptual consistency                 | Logical consistency |
| 4  | domain consistency                     |                     |
| 5  | format consistency                     |                     |
| 6  | topological consistency                |                     |
| 7  | absolute or external accuracy          | Positional accuracy |
| 8  | relative or internal accuracy          |                     |
| 9  | gridded data positional accuracy       |                     |
| 10 | classification correctness             | Thematic accuracy   |
| 11 | non-quantitative attribute correctness |                     |
| 12 | quantitative attribute accuracy        |                     |
| 13 | accuracy of a time measurement         | Temporal quality    |
| 14 | temporal consistency                   |                     |
| 15 | temporal validity                      |                     |

## 2.1 Completeness

### commission

excess data present in a dataset

### omission

data absent from a dataset

## 2.2 Logical Consistency

### conceptual consistency

adherence to rules of the conceptual schema

### domain consistency

adherence of values to the value domains

### format consistency

degree to which data is stored in accordance with the physical structure of the dataset

### topological consistency

correctness of the explicitly encoded topological characteristics of a dataset

## 2.3 Positional Accuracy

### absolute or external accuracy

closeness of reported coordinate values to values accepted as or being true

### relative or internal accuracy

closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true

### gridded data positional accuracy

closeness of gridded data spatial position values to values accepted as or being true

## 2.4 Thematic Accuracy

### classification correctness

comparison of the classes assigned to features or their attributes to a universe of discourse (e.g. ground truth or reference data)

### non-quantitative attribute correctness

measure of whether a non-quantitative attribute is correct or incorrect

### quantitative attribute accuracy

closeness of the value of a quantitative attribute to a value accepted as or known to be true

## 2.5 Temporal Quality

accuracy of a time measurement

closeness of reported time

measurements to values accepted as  
or known to be true

temporal consistency

correctness of the order of events

temporal validity

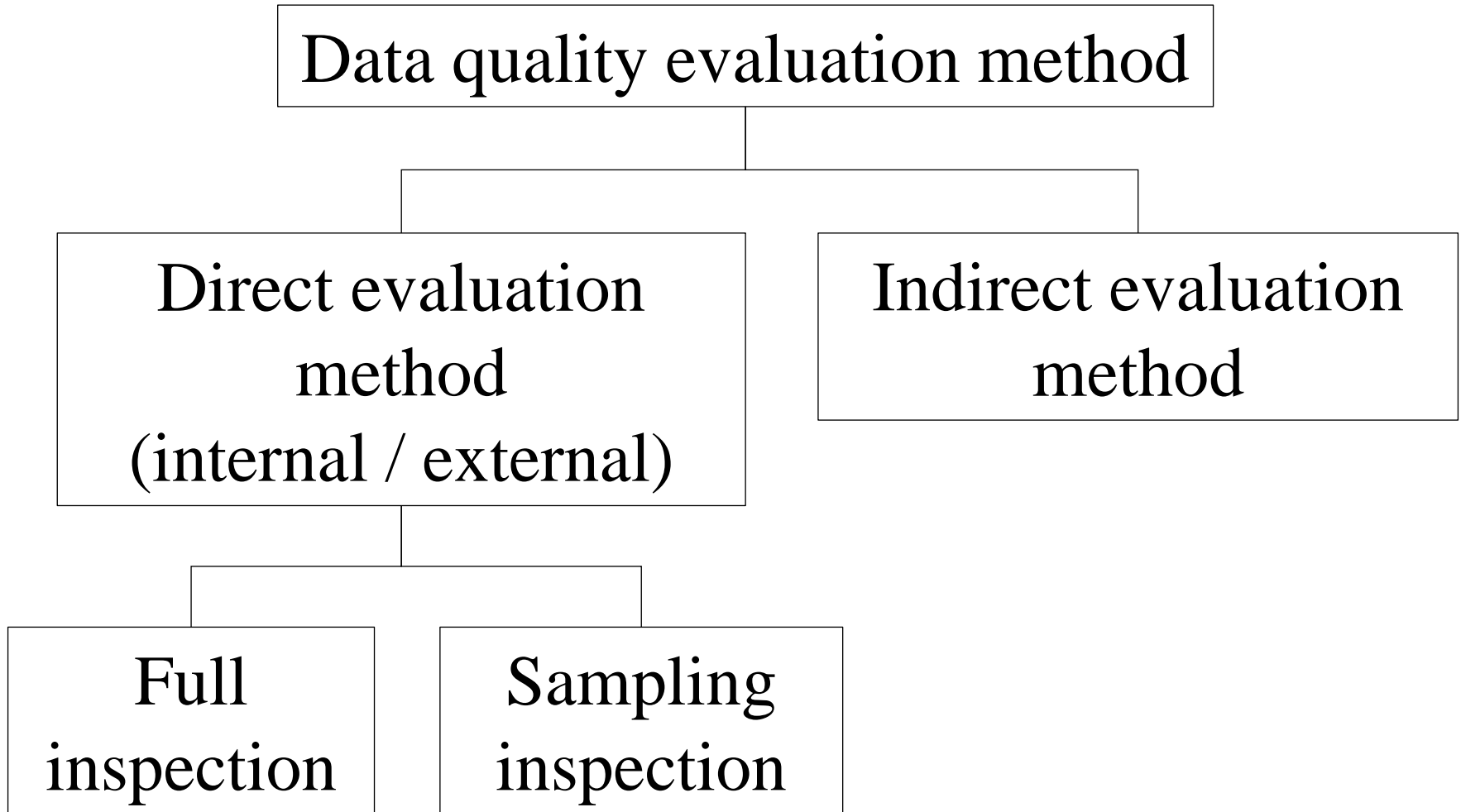
validity of data with respect to time



### 3. Non-quantitative data quality elements

| # | Elements | Definitions  |
|---|----------|--|
| 1 | purpose  | the rationale for creating a data set and contains information about its intended use, which may not be the same as the actual use of the data set         |
| 2 | usage    | the application for which a data set has been used, either by the data producer or by other data users.  |
| 3 | lineage  | the history of a data set and recounts the life cycle of a data set from collection and acquisition through compilation and derivation to its current form |

## 4. Data quality evaluation methods



## 4.1 Direct Evaluation Method

### 1. Internal or External

Internal evaluation: uses only data that resides in the dataset

or

External evaluation: requires reference data (accepted as representing the true values) external to the dataset

### 2. Full Inspection or Sampling

Full Inspection: appropriate for small populations or by automated means

or

Sampling: tests on subsets of the data

## 4.2 Indirect Evaluation Method

based on;

external knowledge or experience of the data product

may include;

Non-quantitative quality information;

usage, lineage and purpose

Data quality reports on the data

Original data used to produce the data

# 5. Reporting of data quality results

Reporting contents of data quality results as metadata

| # | Reporting contents             | Example   |
|---|--------------------------------|---|
| 1 | name of data quality element   | omission (absent item)  |
| 2 | data quality scope             | hospitals in Bangladesh, as of 1st June 2020  |
| 3 | data quality measure           | Number of missing items compared to the original source (hospital list, as of 1st June 2020, by Ministry of Health) |
| 4 | conformance quality level      | 0 = pass, 1 or more = fail (no missing)   |
| 5 | data quality evaluation method | Count the number of missing items using a software function   |
| 6 | data quality result            | 0 (pass)  |
| 7 | data quality value type        | Integer   |
| 8 | data quality value unit        | -   |
| 9 | data quality date              | 2020/06//01   |