

PENILAIAN KUALITI DATA GEOSPATIAL

**Cawangan Pembangunan Polisi, Standard & Outreach
Seksyen Pembangunan Standard
Pusat Geospasial Negara (PGN)**

Apa itu Kualiti Data?

- Data quality is the degree of data excellency that satisfy the given objective. In other words, completeness of attributes in order to achieve the given task can be termed as Data Quality.
- Kualiti data adalah tahap kecemerlangan data yang memenuhi objektif yang diberikan. Dengan kata lain, kelengkapan atribut untuk mencapai tugas yang diberikan dapat disebut sebagai Kualiti Data.
- Data quality is a pillar in any GIS implementation and application as reliable data are indispensable to allow the user obtaining meaningful results.
- Kualiti data adalah tonggak dalam pelaksanaan dan aplikasi GIS kerana data yang dapat dipercayai sangat diperlukan untuk membolehkan pengguna memperoleh hasil yang bermakna.

Skop

- Kualiti data sentiasa dikaitkan dengan ketepatan (accuracy) dan kejituan (precision) dari segi kedudukan.
- Ia merupakan prosedur yang jelas dan konsisten bagi memastikan sejauh manakah sesuatu produk geospasial itu memenuhi kriteria dan spesifikasi yang telah ditetapkan.
- Penetapan prosedur untuk melaksanakan kerja-kerja penilaian kualiti terhadap data geospasial.
- Untuk memastikan kesahihan butiran dan atribut.

Pematuhan

- 3 peringkat pematuhan:
 - Prosedur penilaian kualiti data
 - Langkah-langkah penilaian kualiti data
 - Penyediaan laporan kualiti data

Rujukan

- MS ISO 19101:2003, Geographic Information – Reference model
- MS ISO 19105:2003, Geographic Information – Conformance and testing
- MS ISO 19108:2003, Geographic Information – Temporal schema
- MS ISO 19113:2003, Geographic Information – Quality principles
- MS ISO 19114:2006, Geographic Information – Quality evaluation procedures
- MS ISO 19115:2003, Geographic Information – Metadata referencing by coordinates
- MS ISO 19131:2009, Geographic Information – Data products specifications
- MS 1759:2004, Geographic Information/Geomatics - feature and attribute codes

Istilah dan Definisi

Atribut

- Ciri-ciri bagi sesuatu butiran.
- Contoh: Nilai ketinggian, lebar, panjang

Butiran

- Gambaran objek di bumi sama ada objek semulajadi atau buatan manusia.

Bukan kuantitatif

- Dataset yang tidak boleh dihitung dan dikumpulkan mengikut kumpulan atau atribut.

Dataset

- Kumpulan data yang boleh dikenalpasti sama ada dataset statik atau dinamik.

Elemen kualiti data

- Komponen kuantitatif yang menjelaskan kualiti dataset yang bergantung kepada kandungan dataset dan spesifikasi produk.

Item

- Satu unit minimum untuk diperiksa dan didefinisikan oleh pembekal data berdasarkan kepada spesifikasi produk.

Jenis kualiti data

- Jenis nilai untuk melaporkan keputusan kualiti data seperti Boolean, perausan dan nombor.

Keputusan kualiti data

- Nilai atau satu set nilai yang diperolehi daripada:
 - Ukuran kualiti data
 - Hasil penilaian daripada perolehan sesuatu nilai
 - Set nilai data berbanding dengan tahap pematuhan kualiti yang ditetapkan

Kualiti

- Ciri keseluruhan produk yang mampu memenuhi kehendak yang nyata atau tersirat.

Kuantitatif

- Dataset yang boleh dihitung.

Lot pensampelan

- Satu unit minimum sampel daripada koleksi item di dalam skop kualiti data yang dipilih dan diperiksa. Setiap lot pensampelan seboleh-bolehnya mengandungi item yang mempunyai syarat dan masa yang sama.

Metadata

- Data mengenai data.

Pematuhan

- Memenuhi keperluan spesifikasi yang ditetapkan.

Istilah dan Definisi (samb.)

Pematuhan tahap kualiti

- Nilai threshold atau satu set nilai threshold bagi keputusan kualiti data yang digunakan untuk menentukan sejauh mana dataset memenuhi kriteria yang ditetapkan dalam spesifikasi produk atau keperluan pengguna.

Populasi

- Penjumlahan item-item yang diberi pertimbangan.

Prosedur penilaian kualiti data

- Tatacara yang digunakan dalam membuat penilaian dan melaporkan kaedah penilaian kualiti serta keputusan kualiti berkenaan.

Siri dataset

- Kumpulan dataset yang mempunyai spesifikasi produk yang sama.

Skop kualiti data

- Liputan atau ciri-ciri yang mana maklumat kualiti dilaporkan.

Spesifikasi produk

- Penerangan mengenai universe of discourse dan spesifikasi bagi memetakannya kepada dataset.

Sub-elemen kualiti data

- Komponen bagi elemen kualiti data.

Tarikh kualiti data

- Tarikh atau julat tarikh sesuatu ukuran kualiti data yang digunakan.

Threshold

- Satu tahap yang munasabah merujuk kepada spesifikasi yang ditetapkan.

Ukuran kualiti data

- Penilaian bagi sub-elemen kualiti data.

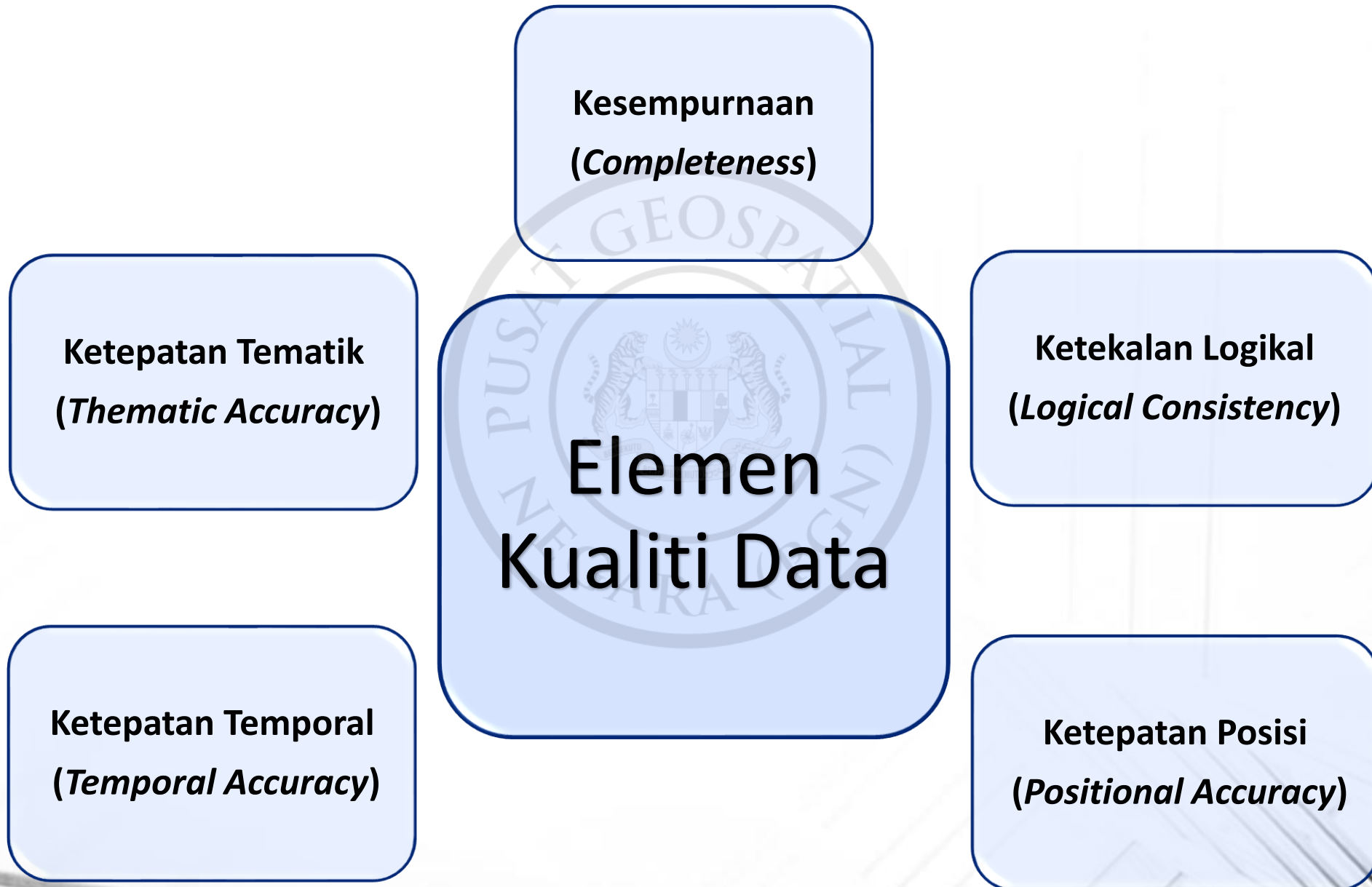
Unit nilai kualiti data

- Unit nilai untuk melaporkan keputusan kualiti data.
- Contoh: meter

Universe of discourse

- Gambaran dunia sebenar atau dunia andaian yang berkepentingan.

Elemen Kualiti Data



Elemen & Sub-Elemen Kualiti Data

Kesempurnaan (*Completeness*)

Kewujudan
(lebih data)

Ketidakwujudan
(kekurangan data)

Ketekalan Logikal (*Logical Consistency*)

Ketekalan konsep
(*conceptual consistency*)

Ketekalan domain
(*domain consistency*)

Ketekalan format
(*format consistency*)

Ketekalan topologi
(*topological consistency*)

Ketepatan Posisi (*Positional Accuracy*)

Ketepatan mutlak/luaran
(*absolute/external accuracy*)

Ketepatan relatif/dalaman
(*relative/internal accuracy*)

Ketepatan posisi data bergrid
(*gridded data position accuracy*)

Ketepatan Temporal (*Temporal Accuracy*)

Ketepatan pengukuran masa
(*accuracy of a time measurement*)

Ketekalan temporal
(*temporal consistency*)

Kesahihan temporal
(*temporal validity*)

Ketepatan Tematik (*Thematic Accuracy*)

Kebenaran klasifikasi
(*classification correctness*)

Kebenaran atribut bukan kuantitatif
(*non-quantitative correctness*)

Ketepatan atribut kuantitatif
(*quantitative attribute accuracy*)

Elemen & Sub-Elemen Kualiti Data

1

**Kesempurnaan
(*Completeness*)**

Kewujudan dan ketidakwujudan butiran, atribut dan hubungannya.

**Kewujudan
(lebih data)**

Wujud lebih dalam dataset.

**Ketidakwujudan
(kekurangan data)**

Wujud kekurangan/ketiadaan dalam dataset.

2

Ketekalan Logikal
(Logical Consistency)

Tahap pematuhan kepada peraturan logical (*logical rules*) sesuatu struktur data, atribut dan hubungannya.

Ketekalan konsep
(conceptual consistency)

Pematuhan/pertalian kepada *conceptual schema*.

Ketekalan domain
(domain consistency)

Pematuhan/pertalian kepada nilai domain.

Ketekalan format
(format consistency)

Pematuhan/penyesuaian penyimpanan data dengan struktur fizikal dataset.

Ketekalan topologi
(topology consistency)

Kebenaran terhadap topologi yang dikodkan (*encoded topology*) khusus bagi dataset.

3

Ketepatan Posisi
(Positional Accuracy)

Ketepatan bagi posisi sesuatu butiran.

Ketepatan mutlak atau luaran
(absolute or external accuracy)

Tahap kehampiran/ketepatan koordinat yang dinyatakan berbanding dengan nilai benar yang diterima pakai.

Ketepatan relatif atau dalaman
(relative or internal accuracy)

Tahap kehampiran/ketepatan koordinat bagi kedudukan relatif butiran berbanding dengan kedudukan relatif yang diterima pakai.

Ketepatan posisi data bergrid
(gridded data position accuracy)

Tahap kehampiran/ketepatan koordinat bagi nilai kedudukan bergrid berbanding dengan nilai benar yang diterima pakai.

4

Ketepatan Temporal
(Temporal Accuracy)

Ketepatan tempoh masa bagi atribut dan tempoh hubungan antara butiran.

Ketepatan pengukuran masa
(accuracy of time measurement)

Ketepatan rujukan temporal bagi sesuatu item.

Ketekalan temporal
(temporal consistency)

Ketepatan turutan kejadian/peristiwa, sekiranya dilaporkan.

Kesahihan temporal
(temporal validity)

Tempoh sah data yang digunakan merujuk kepada masa.

5

Ketepatan Tematik
(Thematic Accuracy)

Ketepatan bagi atribut kuantitatif, kebenaran (correctness) bagi atribut bukan kuantitatif dan pengkelasan butiran serta hubungannya.

Kebenaran klasifikasi
(classification correctness)

Perbandingan terhadap klasifikasi yang diberi kepada sesuatu butiran atau atribut bagi *universe of discourse* (dataset berkenaan).

Kebenaran atribut bukan kuantitatif
(non-quantitative attribute correctness)

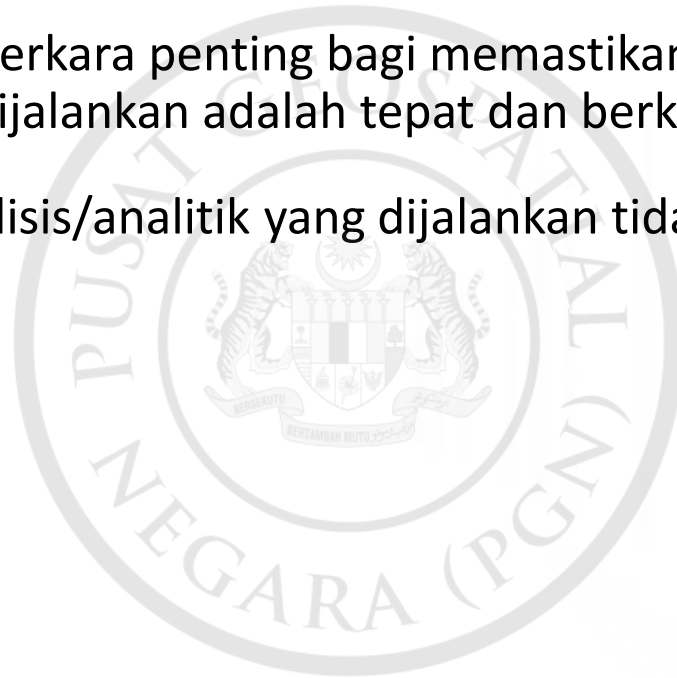
Ketepatan bagi atribut bukan kuantitatif.

Kebenaran atribut kuantitatif
(quantitative attribute correctness)

Ketepatan bagi atribut kuantitatif.

Gambaran Keseluruhan Elemen Kualiti Data

- Kualiti data geospasial satu perkara penting bagi memastikan produk akhir/analisis/analitik yang dijalankan adalah tepat dan berkesan.
- Tanpa pernyataan kualiti, analisis/analitik yang dijalankan tidak dapat kesahihan.



Hubungkait Antara Elemen

- Elemen-elemen kualiti data mempunyai hubungkait antara satu sama lain.
- Hubungan tersebut melibatkan dataset, butiran dan atribut bagi sesuatu produk.
- Penilaian kualiti data bermula dari elemen kesempurnaan, diikuti sama ada dengan ketekalan logik, ketepatan posisi, ketepatan temporal dan ketepatan tematik.
- Contoh: satu dataset (peta topografi) yang mengandungi pelbagai butiran (seperti tanaman, jalan, bangunan dll.) dan atribut (jenis, bilangan, kod, dll.), setelah penilaian dijalankan dan didapati tidak mematuhi elemen kesempurnaan (tidak mempunyai butiran tanaman). Maka ujian ke atas dataset ini tidak dapat diteruskan.

Proses Penilaian Kualiti Data

Kenalpasti elemen berkaitan, sub-elemen dan skop kualiti data

- Langkah ini dijalankan berulang kali mengikut kekerapan penilaian yang dijalankan.

Kenalpasti ukuran kualiti data

- Langkah ini perlu dijalankan bagi setiap ujian yang dijalankan.

Pilih dan gunakan kaedah penilaian kualiti data

- Langkah ini diperlukan untuk menentukan kaedah penilaian langsung atau tidak langsung.

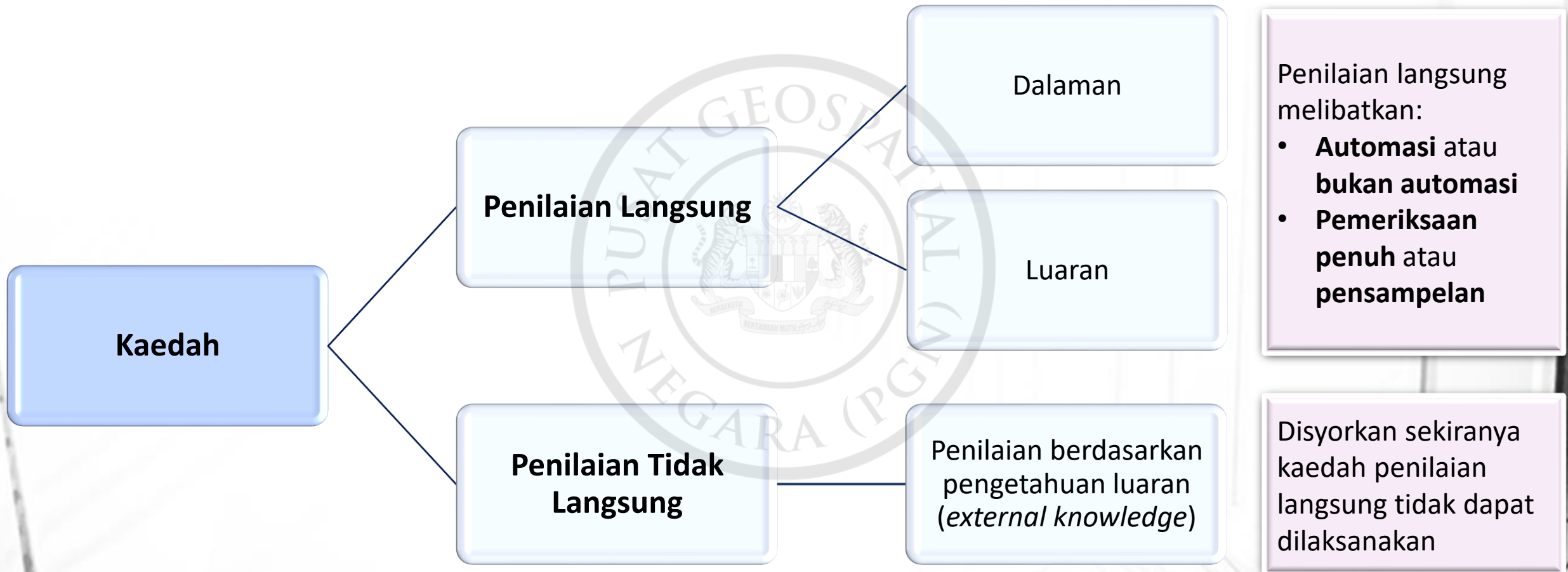
Tentukan hasil kualiti data

- Keputusan kualiti data kuantitatif, nilai kualiti data atau set nilai kualiti data, unit nilai kualiti dan tarikh.

Tentukan pematuhan

- Tahap pematuhan perlu dibandingkan dengan hasil kualiti data bagi menentukan tahap pematuhan (lulus atau gagal).

Kaedah Penilaian Kualiti Data

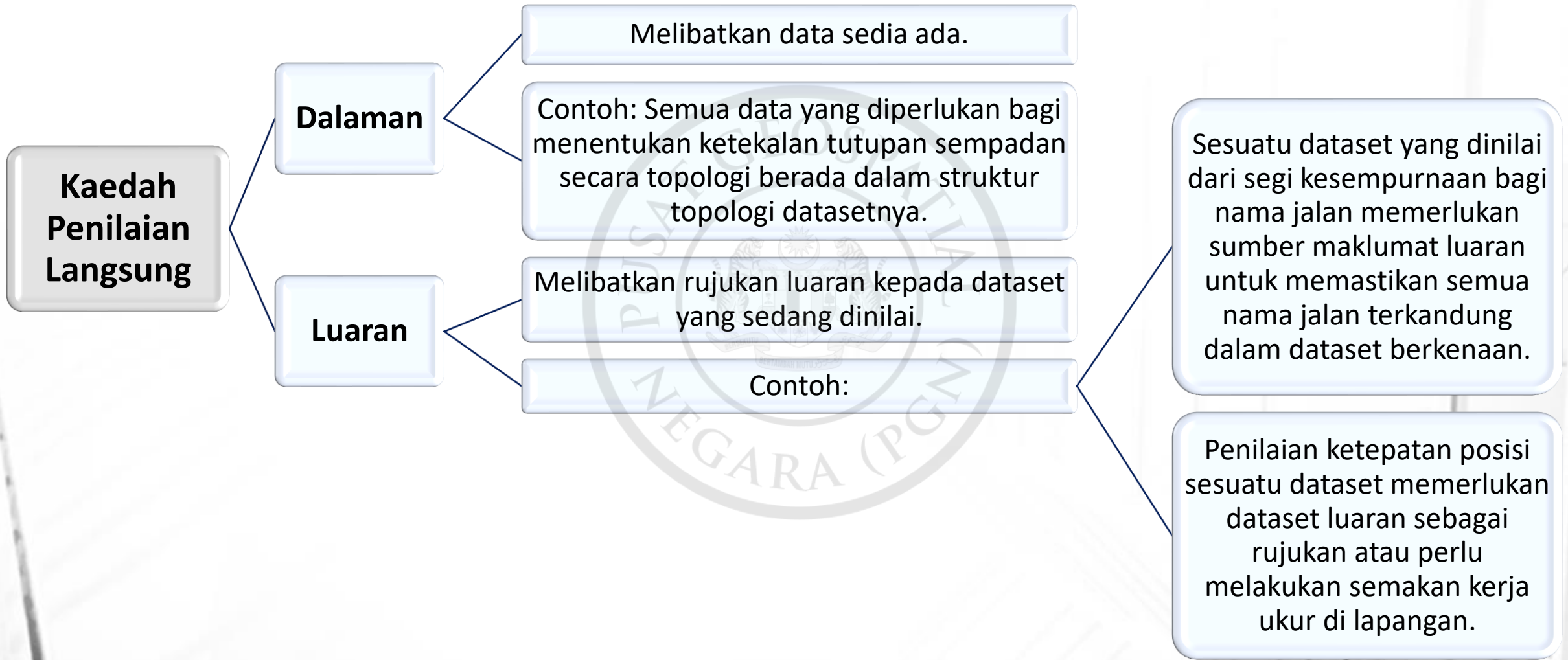


Penilaian langsung melibatkan:

- **Automasi** atau **bukan automasi**
- **Pemeriksaan penuh** atau **pensampelan**

Disyorkan sekiranya kaedah penilaian langsung tidak dapat dilaksanakan

Kaedah Penilaian Kualiti Data (samb.)



Kesimpulan

- Kualiti data geospasial satu perkara penting bagi memastikan produk akhir/analisis/analitik yang dijalankan adalah tepat dan berkesan.
- Tanpa pernyataan kualiti, analisis/analitik yang dijalankan tidak dapat kesahihan.

