

UWI1B2 LITERASI DATA

Kualitas Data

Anisa Herdiani, S.T., M.T.

Ir. Candiwan, M. ICT

Dr. Doan Perdana, S.T., M.T.




Capaian Pembelajaran

Mahasiswa mampu menjelaskan faktor-faktor penentu kualitas data



Kualitas Data

Kualitas data merupakan kondisi data yang sesuai dengan kebutuhan penggunaanya






Kapan kita harus mengevaluasi kualitas dari data yang kita miliki?






Mengevaluasi data menggunakan intuisi

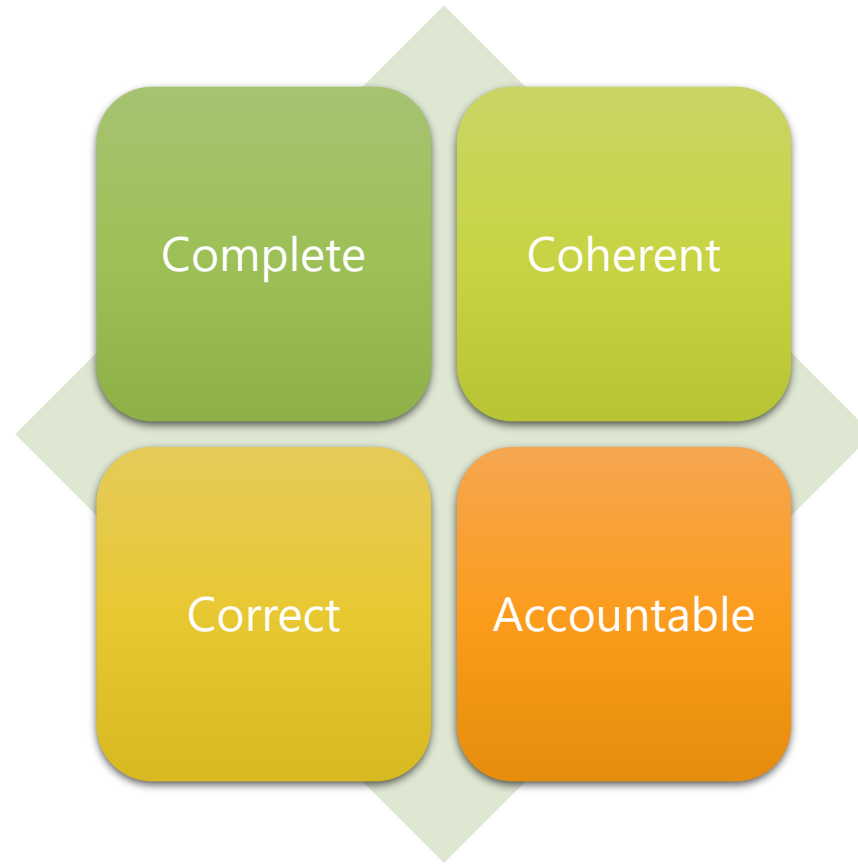
- Tidak dapat mengidentifikasi hal yang samar
 - Bisa jadi salah
 - Sangat bergantung kepada pengalaman yang dimiliki
- 



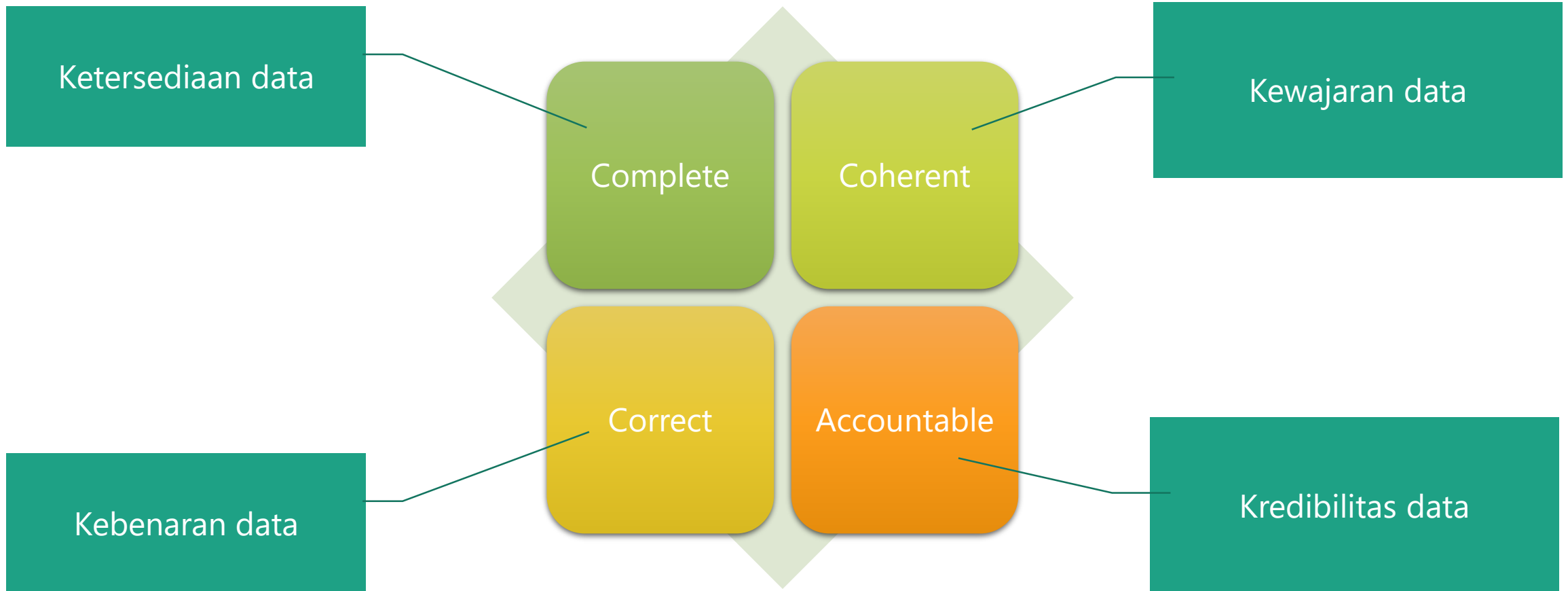
Framework Kualitas Data

- Panduan yang efisien dan konsisten sehingga kita bisa berada pada tingkatan pemahaman yang lebih baik terhadap karakteristik dan kualitas dari data yang kita miliki
- 

Framework Kualitas Data



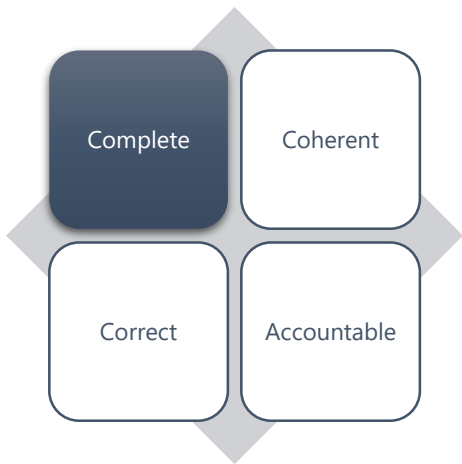
Framework Kualitas Data



Complete

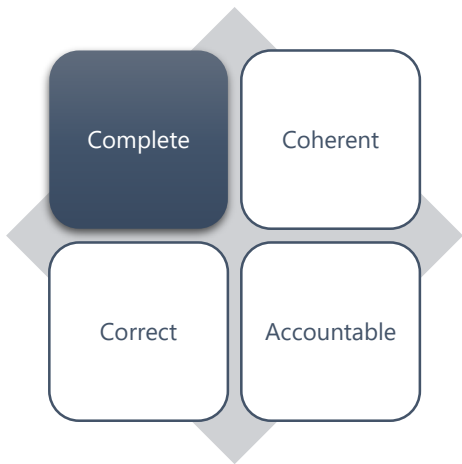
Apakah saya memiliki semua data yang saya butuhkan untuk menjawab pertanyaan saya?


1. Tersedianya semua variabel yang dibutuhkan
2. Tersedianya semua unit observasi yang dibutuhkan
3. Tersedianya semua nilai dari variabel pada unit observasi



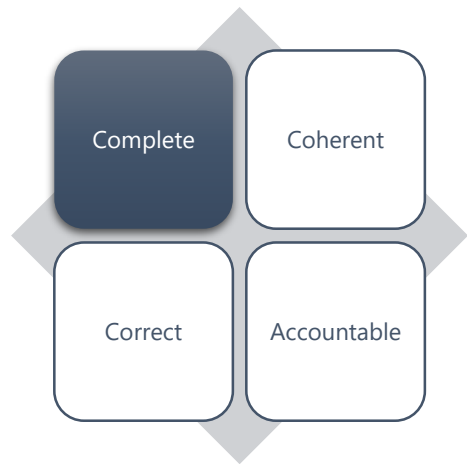


Kelengkapan Jumlah Record



- Apakah sebenarnya Anda memiliki semua record untuk pertanyaan yang diajukan?
 - Pastikan jumlah data yang dimiliki sesuai dengan data di system lainnya
- 

Kelengkapan Jumlah Record

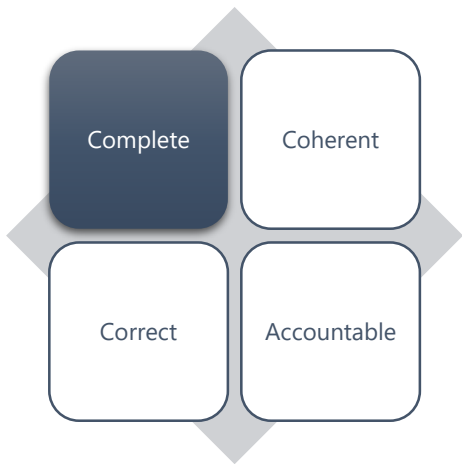


- Untuk data yang jumlahnya sangat besar, terutama karena sekarang kita berada di era Big Data, pengecekan data dapat dilakukan menggunakan teknik sampling.

Record-Level Completeness

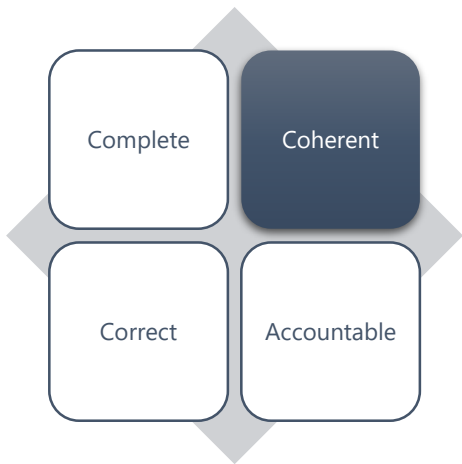
Ketika menemukan ketidaklengkapan data pada level baris, pilihan aksinya adalah:

1. Perbaiki data yang hilang.
2. Hapus baris yang tidak lengkap.
3. Memberikan tanda untuk baris yang tidak lengkap dan menginformasikan adanya ketidaklengkapan data ketika menjawab kebutuhan atas suatu informasi.
4. Tidak melakukan apa-apa.



Coherent

Apakah data yang dimiliki konsisten dan sesuai dengan logika?

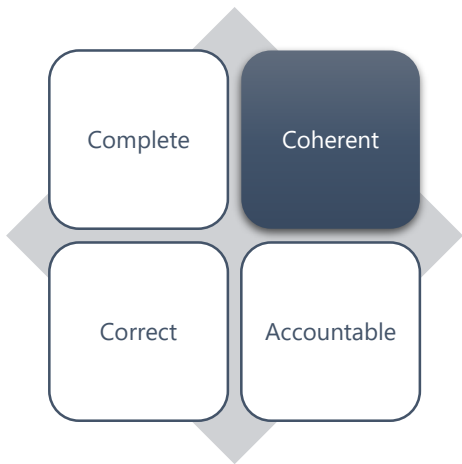


Referential integrity

Referential integrity adalah sebuah cara untuk menjaga konsistensi data antara tabel yang saling berhubungan. Cara ini diterapkan untuk data yang disimpan dalam bentuk tabular.

Contoh:

Terdapat pengecekan ke dalam data mata kuliah tawar program studi sedemikian sehingga hanya mata kuliah yang ditawarkan saja yang dapat diambil oleh mahasiswa ketika masa registrasi semester baru.



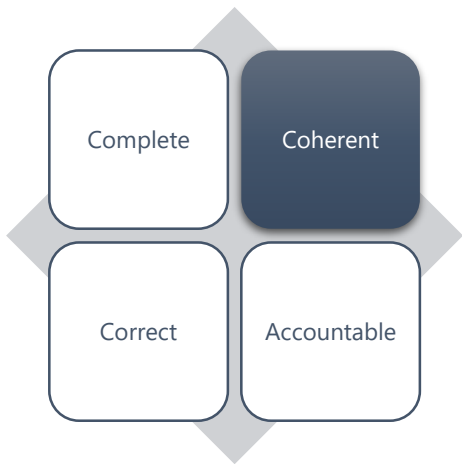
Value Integrity

Apakah nilai yang disimpan konsisten secara internal?

Contoh:

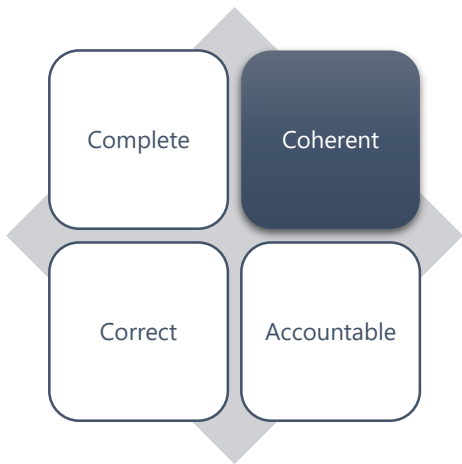
Kuota yang tersedia untuk suatu mata kuliah X adalah 40 mahasiswa. Jika sudah ada 30 mahasiswa yang melakukan registrasi dan mengambil mata kuliah X, maka kuota yang tersisa seharusnya adalah 10 mahasiswa.

Informasi yang terdapat pada data registrasi, data mata kuliah (yang mencakup kuota mahasiswa) harus dipastikan konsistensinya agar jumlah pendaftar tidak melebihi kuota yang tersedia.



Evaluasi Koherensi

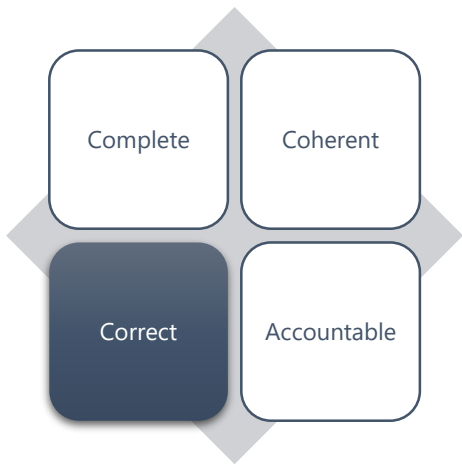
- Tentukan level koherensi yang dibutuhkan.
 - Apakah validasi pada referential integrity atau value integrity, atau mungkin diperlukan validasi integritas data yang lebih kompleks?
- Tentukan seberapa lengkap validasi yang dibutuhkan, dan apa saja batasan performansi dan waktunya.
 - Apakah perlu melakukan validasi untuk setiap baris atau setiap relasi antar table yang ada?
 - Ataukah dapat dilakukan *statistical sampling* untuk mengevaluasi sebagian baris-baris yang representatif?





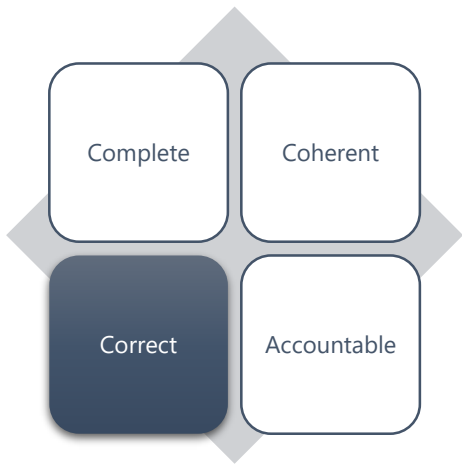
Correct

Apakah data Anda sudah benar?



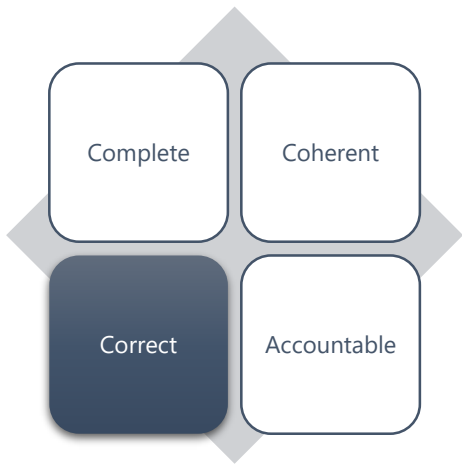
Evaluasi Kebenaran

- Rincilah data apa saja yang dapat diverifikasi dengan mudah.
 - Misal Anda memiliki data penjualan dari sebuah supermarket, apakah data yang perlu Anda analisis seharusnya adalah data penjualan yang terjadi pada tahun 2018-2019?
- Tentukan hal apa saja yang penting untuk divalidasi
- Pahami, seberapa banyak data yang harus benar?



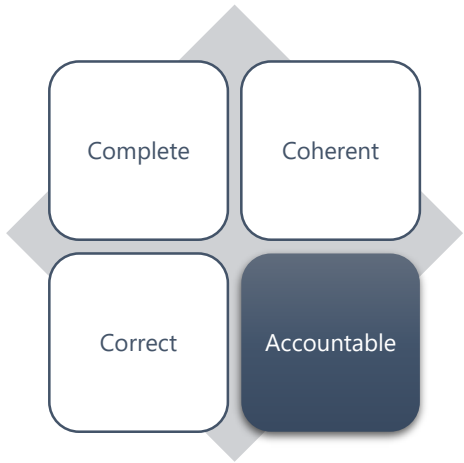
Evaluasi Kebenaran

- Putuskan apa yang akan dilakukan pada data yang tidak benar
 - Apakah mungkin untuk memperbaiki datanya?
 - Bisakah proses dilanjutkan tanpa menggunakan data yang tidak benar tersebut?
 - Beri tanda untuk data yang tidak benar, dan nyatakan fakta tersebut dalam proses analisis.
 - Pilihan lainnya, pisahkan data yang tidak benar, dan lakukan analisis secara terpisah.

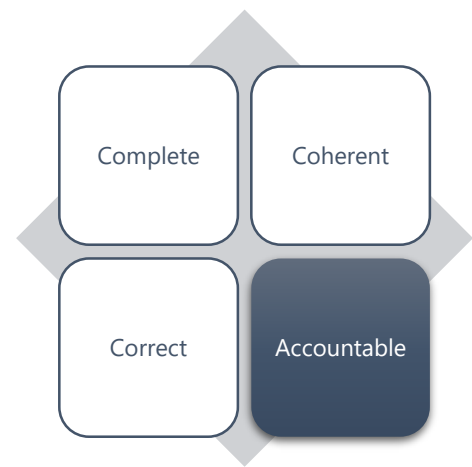


Accountability

Siapa yang bertanggung jawab terhadap data Anda?



Data Flow



Mendapatkan data dari sumber tertentu

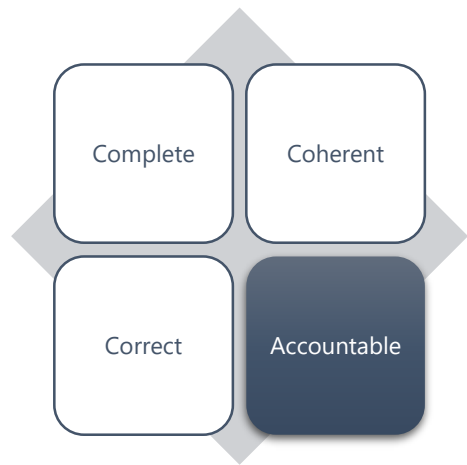
Modify

Melakukan pembersihan data, memperkaya data, atau mengambil sebagian data untuk tujuan tertentu



Menggunakan data untuk mengarahkan keputusan internal untuk kemudian didistribusikan ke pihak eksternal

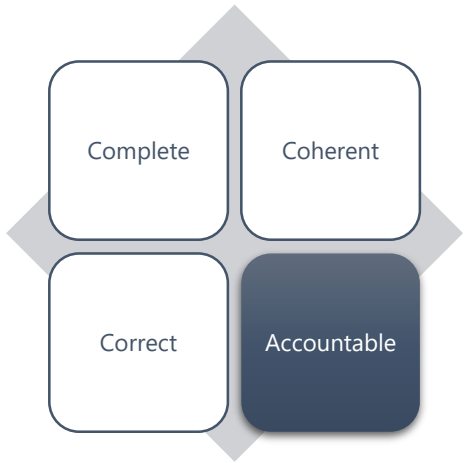
Catat Semua Sumber Data



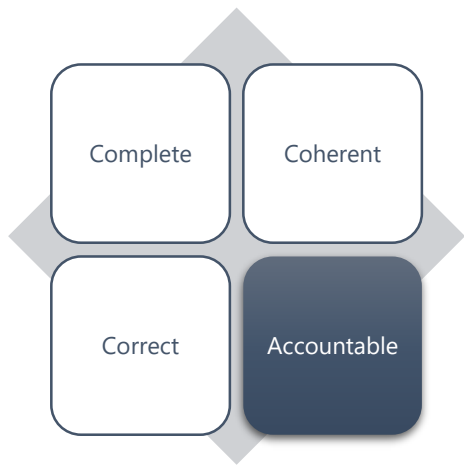
- Bagaimana cara mengakses datanya?
- Tentukan pihak mana yang memiliki dan mengelola setiap data yang didapatkan
- Tentukan siapa saja yang dapat dihubungi jika ada pertanyaan terkait data dan apa saja tanggung jawab atau kesepakatan layanan dari datanya : Apa jaminannya terkait kualitas, akurasi, format data, dan ketepatan waktu penyediaan datanya.

Simpan semuanya

- Simpan data asli (data mentah) yang didapatkan.
- Data asli dapat digunakan untuk mengecek modifikasi apa yang sudah dilakukan

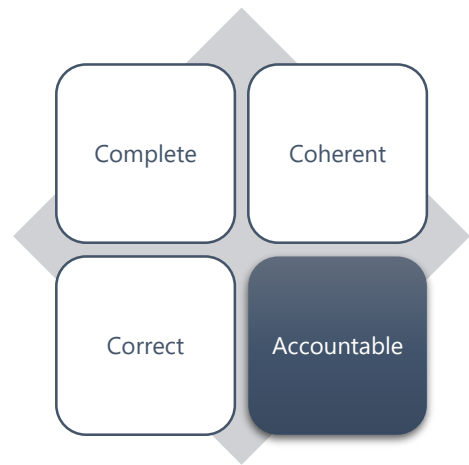


Lakukan Audit Mandiri



- Periksa data untuk mengkonfirmasi apakah salinan data yang dimiliki sesuai dengan versi yang berada di sumber data.
- Konfirmasi bahwa setiap modifikasi atau peningkatan kualitas data tidak memunculkan kesalahan pada data.
- Pastikan setiap data tidak mengalami modifikasi tanpa rencana atau berubah untuk tujuan tertentu yang tidak sesuai (dipalsukan).

Telusuri Akses terhadap Data



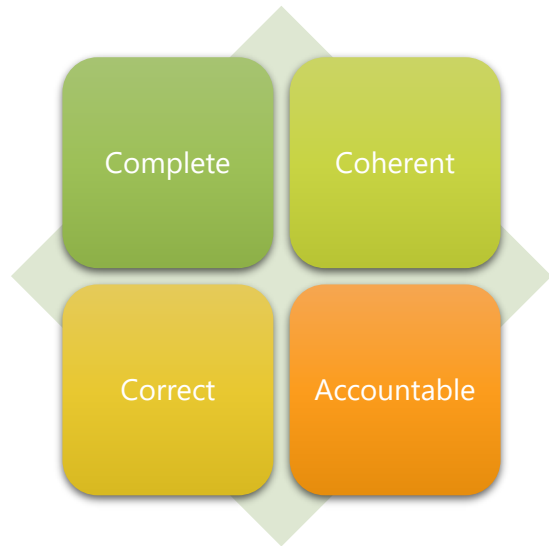
- Pahami siapa saja yang dapat mengakses data dan bagaimana caranya
- Untuk data yang tersimpan dalam sistem pengelolaan data yang mapan, pengaksesan data dapat dikendalikan dan dimonitor dengan baik hingga pada level apa saja data yang dapat diakses oleh setiap pengguna.

Kesimpulan



- Memperhatikan kualitas data sebelum mulai bekerja dengan data tidak hanya menghemat waktu, melainkan juga memberikan pemahaman yang lebih baik mengenai apa saja kemampuan dan keterbatasan Anda dalam menganalisis data.
- Pastikan Anda membuat dan mendokumentasikan keputusan dan aksi yang dilakukan terkait perhatian Anda terhadap kualitas data.

Kesimpulan



- Analisis terhadap kualitas data dan penyiapan data adalah langkah awal yang penting untuk mendapatkan *insight* dari data yang dimiliki.



Referensi

- McCallum, Q. Ethan. 2013. *Bad Data Handbook. Mapping the World of Data Problems*. Sebastopool, CA: O'Reilly Media, Inc.

