

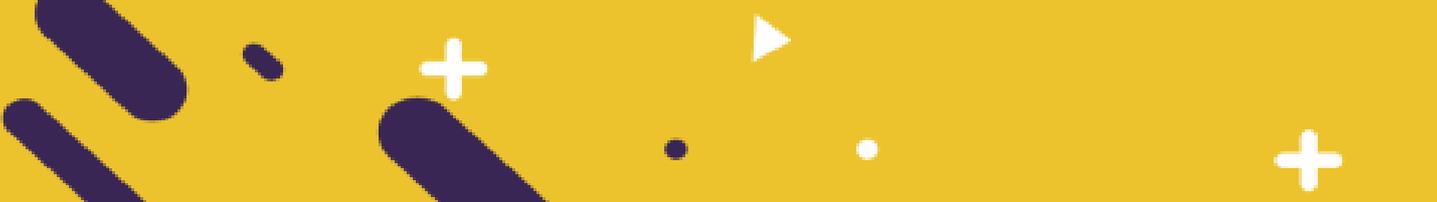


# PETUNJUK TEKNIS

Mempersiapkan  
Data Terbuka



DINAS KOMUNIKASI DAN INFORMATIKA  
KABUPATEN BANTUL



VOLUME 1

# Petunjuk Teknis

# Mempersiapkan

# Data Terbuka

## Pendahuluan

Bagian ini berisi beberapa teknik dan standar yang sangat disarankan untuk digunakan ketika melakukan standardisasi data menjadi lebih *machine-readable* dan *reusable*. Sebagai informasi, bagian ini merupakan saduran dari dokumen **“Pedoman Umum Data Terbuka: Panduan Penyusunan Data Terbuka Pemerintah DKI Jakarta”** yang disusun bersama-sama oleh berbagai pihak dengan Open Data Lab Jakarta sebagai koordinator utama. Tautan publik dokumen tersebut dapat dirujuk di: <http://opendatalabjakarta.github.io/bp-penyusunan/kandidat/index.html>

Lampiran ini berisi informasi bagaimana cara menyiapkan metadata, melakukan Penyusunan Data, dan memberikan petunjuk bagaimana cara melakukan Penulisan Data yang baik.

# Metadata

Metadata dapat dipahami sebagai keterangan mengenai data atau seringkali disebut sebagai data mengenai data. Di dalam Data.go.id ini, sistem metadata yang dirancang terdiri atas beberapa field yang wajib diisi. Tujuan dari adanya sistem metadata ini adalah untuk proses dokumentasi dan pengarsipan dari seluruh data yang ada di dalam Data.go.id.

Daftar metadata yang perlu diisi di dalam Data.go.id adalah sebagai berikut:

NAMA FIELD	KETERANGAN FIELD	CONTOH PENGISIAN	WAJIB/TIDAK WAJIB
Judul	Judul dari kumpulan data. Haru singkat, dan jelas	"Anggaran Pendapatan dan Belanja Negara"	WAJIB
Deskripsi	Mudah dimengerti dan menjelaskan hal-hal dan catatan penting dalam data set	"Kumpulan data berisi mengenai jumlah pendapatan dan belanja Republik Indonesia yang dianggarkan"	WAJIB
Tag	Kata Kunci untuk memudahkan pencari kumpulan data	"anggaran", "pemerintah daerah", "keuangan", "pendapatan", "belanja"	WAJIB
Lisensi	Lisensi yang digunakan oleh kumpulan data	"Creative Commons Attribution"	WAJIB
Organisasi	Organisasi/Institusi yang menerbitkan kumpulan data	"Kementerian Keuangan"	WAJIB
Visibility	Yang bisa melihat kumpulan data ini	"Publik"	WAJIB

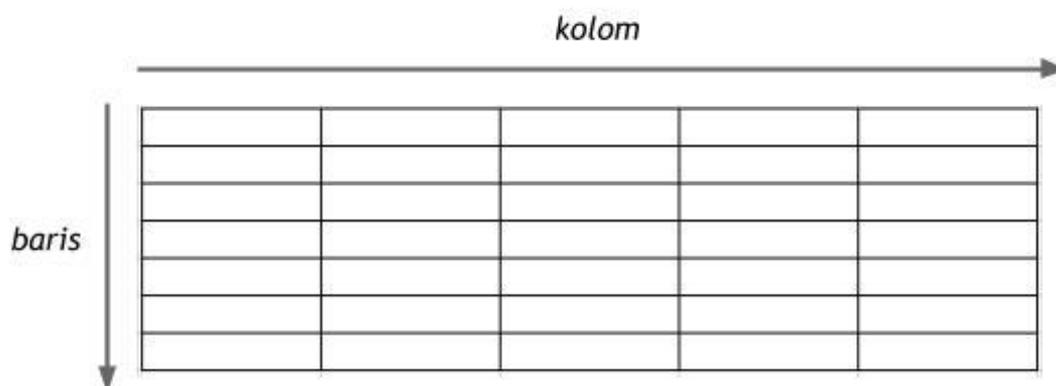
File	File berisi data	"APBN_2014.csv" atau "www.bojonegoro.go.id/apbn.csv"	WAJIB
Name	Nama dari data	"APBN 2014"	WAJIB
Description	Deskripsi singkat atau keterangan tambahan mengenai data tersebut	"APBN Indonesia untuk tahun 2014"	TIDAK WAJIB
Format	Format dari data	"CSV"	WAJIB
Sumber	Sumber dari mana data tersebut diambil (kalau berasal dari situs di luar portal data)	"www.kemenkeu.go.id/bankdata"	TIDAK WAJIB
Nama Kontak	Nama kontak yang bertanggung jawab terhadap data set ini	"Prasetya Dwicahya"	WAJIB
Email Kontak	Email kontak yang bertanggung jawab terhadap data set ini	"prasetya@yahoo.co.id"	WAJIB
Frekuensi Penerbitan	Frekuensi penerbitan dari kumpulan data tersebut. Apakah real-time, satu minggu sekali, satu bulan sekali, dan seterusnya	"Triwulan"	WAJIB
Level Penyajian	Level penyajian dari data yang disajikan di	"Rincian Objek"	WAJIB

	dalam Portal		
Tahun	Berisi interval waktu atau tahun atau bulan atau periode yang mendeskripsi kondisi di mana kumpulan data ini relevan	"2010 – 2014"	WAJIB
Cakupan	Daerah atau wilayah tempat data tersebut berlaku atau diambil	"Nasional"	WAJIB
Rujukan	Berisi tautan yang berisi keterangan lebih lengkap mengenai kumpulan data ini yang biasanya berada pada situs lain di luar situs Data.go.id.	"www.kemkeu.go.id/keteranganapbn.html"	TIDAK WAJIB

Dokumen "Tata Cara Pengisian Metadata" bisa menjadi rujukan lebih lengkap untuk mengetahui bagaimana langkah-langkah untuk melengkapi metadata yang ada di dalam Data.go.id.

## Penyusunan Data

Data perlu disusun secara terstruktur untuk memudahkan analisis dan penggunaan kembali. Bentuk tabular umum dipakai dalam menyusun data secara terstruktur. Bentuk ini memiliki dua dimensi berupa kolom dan baris dimana rentang kolom merepresentasikan jenis pengukuran dan satuan baris merepresentasikan nilai data. Gambar di bawah ini memperlihatkan ilustrasi bentuk tabular.



Dalam pemanfaatan informasi digital perlu diperhatikan pula bahwa bentuk tabular bukanlah satu-satunya bentuk data yang dapat dipakai. Terdapat bentuk data lainnya seperti audio, video, foto, ilustrasi digital dan sebagainya.

Berikut adalah poin-poin penting mengenai penyusunan data tabular secara baik dan benar:

- **Nama-nama kolom harus termuat dalam satu baris**  
Nama-nama kolom umumnya disusun dalam satu baris dan ditempatkan pada baris pertama. Jika nama-nama tersebut tersusun dalam beberapa baris maka perlu diatur ulang agar terangkum dalam satu baris (lihat Sub-bab 3.1).
- **Gunakan satu tipe data per kolom**  
Ketidakjelasan tipe data dalam satu kolom dapat menyebabkan data sulit dianalisis secara otomatis. Oleh karena itu, data yang mengandung banyak informasi perlu dipisahkan dalam beberapa kolom, jika diperlukan. (lihat Sub-bab 3.2).
- **Pastikan pengisian data berbentuk baris, bukan kolom**  
Setiap data baru yang dimasukkan harus membentuk baris baru, bukan kolom baru. Kolom baru hanya boleh ditambahkan apabila pengukuran baru diperkenalkan dalam metode pengumpulan. Jika ditemui bahwa memasukkan data baru berarti selalu memperkenalkan kolom baru maka skema data perlu diatur ulang (lihat Sub-bab 3.3).

- **Pisahkan data mentah dengan data hasil analisis**

Data harus disajikan dalam bentuk paling primer, atau data mentah. Hal ini dikarenakan data mentah dapat dipakai berulang dalam ragam analisis yang berbeda. Data hasil analisis hendaknya disajikan terpisah dari data mentah (lihat Sub-bab 3.4).

Berikut ini akan ditampilkan beberapa contoh kasus umum yang sekiranya dapat membantu pemahaman aturan praktis yang disebutkan di atas.

### **Nama Kolom Bersusun**

Sering kali ditemukan struktur tabular dengan nama-nama kolom tersusun secara hirarkis (lihat ilustrasi). Penyusunan nama kolom seperti ini akan sukar diproses oleh komputer.

Unit	Model A			Model B		
	Panjang	Tebal	Tinggi	Panjang	Tebal	Tinggi
Kotak 1	10	5	8	12	5	10
Kotak 2	15	10	20	15	10	30
Kotak 3	5	5	5	10	10	10

Struktur tabular di atas dapat disusun ulang agar nama-nama kolom berada dalam satu baris:

unit	jenis_model	panjang	tebal	tinggi
Kotak 1	A	10	5	8
Kotak 1	B	12	5	10
Kotak 2	A	15	10	20
Kotak 2	B	15	10	30
Kotak 3	A	5	5	5
Kotak 3	B	10	10	10

### **Nilai Data Majemuk**

Pengisian nilai data dapat melibatkan lebih dari satu elemen informasi. Masing-masing elemen tersebut dapat memiliki tipe data yang berbeda (lihat ilustrasi). Penyajian data seperti demikian perlu diuraikan agar masing-masing elemen dapat dimengerti baik oleh komputer.

Tanggal	Biaya
Selasa, 2 September 2014	Rp 1.200.000,00
Rabu, 3 September 2014	US\$ 100
Kamis, 11 September 2014	US\$ 1,100
Jumat, 12 September 2014	Rp 3.450.000,00

Informasi tanggal dan biaya dalam struktur tabular di atas dapat diurai menjadi sebagai berikut:

tanggal	nilai_biaya	mata_uang
2014-09-02	1200000	IDR
2014-09-03	100	USD
2014-09-11	1100	USD
2014-09-12	3450000	IDR

Perhatikan bahwa nama hari tidak perlu diikutsertakan dalam penyajian informasi tanggal. Hal ini dikarenakan keterangan hari dapat dihasilkan ulang melalui fungsi analisis.

### Tabel dengan Kolom Terus Bertambah

Terkadang banyaknya jenis pengukuran data tidak dapat ditentukan secara pasti dari awal. Hal ini dapat menyebabkan penyusunan skema data yang tidak efisien. Salah satu contoh kasus nyata tersebut adalah bertambahnya jumlah kolom dari waktu ke waktu ketika data baru dimasukkan ke pencatatan (lihat ilustrasi).

Bulan	Meja	Kursi	Papan Tulis	Lemari
Januari	12			
Februari	15	24		
Maret		30	1	
April				3

Tabel di atas memperlihatkan hasil pendataan peralatan secara berkala. Perhatikan bahwa di setiap pergantian bulan selalu terdapat penambahan kolom jenis peralatan yang ingin dicatat. Akibatnya struktur tabular ini dapat mengurangi efisiensi penyimpanan dan pengolahan data. Struktur di atas dapat diperbaiki dengan mengubahnya sebagai berikut:

bulan	jenis_peralatan	jumlah
Januari	Meja	12
Februari	Meja	15
Februari	Kursi	24
Maret	Kursi	30
Maret	Papan Tulis	1
April	Lemari	3

Kasus ini lazim ditemui dalam proses pencatatan data terkait dengan proses memasukkan data observasi atau survei lapangan.

## Reproduksi Data Komputasi

Menyatukan data primer (atau data mentah) dengan data hasil komputasi pada dasarnya tidak direkomendasikan. Hal ini untuk mencegah penyisipan data tidak valid yang dapat dimungkinkan akibat kesalahan teknis perhitungan.

Solusi yang direkomendasikan adalah dengan membuat dua salinan data. Salinan pertama berisikan data primer saja dan salinan kedua berisikan juga nilai-nilai hasil komputasi. Rumus analisis untuk mereproduksi hasil komputasi perlu disertakan dalam deskripsi data agar dapat divalidasi.

Unit	Panjang	Tebal	Tinggi	Volume
Kotak 1	10	5	8	400
Kotak 2	15	10	20	3000
Kotak 3	5	5	5	150

Tabel di atas memiliki kolom Volume yang didapat dari perkalian tiga kolom sebelumnya (*yi*. Panjang, Tebal dan Tinggi). Perhatikan bahwa kesalahan teknis perhitungan mungkin dapat terjadi dan hal ini dapat menyebabkan validitas keseluruhan data menjadi diragukan (*cth*. terdapat kesalahan dalam perhitungan volume Kotak 3).

unit	panjang	tebal	tinggi
Kotak 1	10	5	8
Kotak 2	15	10	20
Kotak 3	5	5	5

Dengan menyajikan salinan data yang berisikan data primer saja, penggunaan data yang tidak valid dapat dihindari.

## Penulisan Data

Bagian ini berisikan pedoman penulisan data secara baku menurut masing-masing tipe data. Penentuan format baku penulisan data sangat penting untuk meningkatkan interoperabilitas data dan keseragaman. Dokumen ini membagi tipe data ke dalam 8 kategori dasar, antara lain:

- Teks,
- Angka,

- Logika Biner,
- Tanggal & Waktu (*sub-tipe*: tanggal, bulan, tahun, tanggal kalender dan waktu),
- Lokasi (*sub-tipe*: garis lintang, garis bujur dan alamat),
- Telepon,
- Email, dan
- URL.

Berikut ini akan dibahas secara detail satu per satu tipe data.

## Teks

Penulisan tipe data Teks memiliki kebebasan penggunaan semua karakter tulisan.

Contoh penulisan benar:

Jakarta Pusat
SND129:21/2014
Rp 20.000,00
Daerah Khusus Ibukota Jakarta (DKI Jakarta, Jakarta Raya) adalah ibu kota negara Indonesia.

Sebagai catatan, ada beberapa teks yang format penulisannya akan dibakukan untuk keseragaman, antara lain penulisan alamat bangunan, nomor telepon, alamat email dan URL situs.

## Angka

Penulisan tipe data Angka terbatas pada karakter numerik [0...9]<sup>1</sup>, karakter penanda positif atau negatif, dan karakter pemisah desimal.

Karakter penanda positif menggunakan tanda plus (+) dan penanda negatif menggunakan tanda minus (-). Tergantung dari konfigurasi lokalisasi di sistem operasi komputer, karakter pemisah desimal dapat berupa tanda titik (.) atau tanda koma (,). Dokumen ini **merekomendasikan** penggunaan tanda titik untuk meningkatkan interoperabilitas penggunaan data angka tersebut.

Contoh penulisan benar:

125
100000
-100000

<sup>1</sup> Penyajian [0...9] dibaca sebagai rentang nilai dari angka 0 sampai angka 9, yi. 0, 1, 2, 3, 4, 5, 6, 7, 8,

123.45
0.000012345

Dalam menuliskan tipe data Angka, karakter-karakter khusus seperti penanda finansial atau persentase harus dihindari. Penggunaan karakter-karakter tambahan ini akan menyebabkan tipe data berubah menjadi tipe data teks. Akibatnya fungsi analisis data dapat mustahil diaplikasikan.

Contoh penulisan salah:

10.000.000
Rp10.000.000,00
2.5%

### Logika Biner

Penulisan tipe data Logika Biner terbatas pada karakter satu ('1') dan nol ('0') yang merepresentasikan dua kondisi: *benar* atau *salah*. Alternatif lain menuliskan tipe data ini adalah dengan menggunakan istilah: "ya" dan "tidak".

Simbol	Arti
1	<i>benar</i>
0	<i>salah</i>
ya	<i>benar</i>
tidak	<i>salah</i>

### Tanggal & Waktu

Penulisan tipe data Tanggal & Waktu berdasarkan kalender Gregorian yang menghitung perubahan waktu menurut gerakan perputaran bumi terhadap matahari. Tipe data ini terbagi menjadi 5 sub-tipe sesuai detail kebutuhan, antara lain:

- tanggal
- bulan
- tahun
- tanggal kalender
- waktu

#### Tanggal

Penulisan tipe data Hari terbatas pada karakter 2-angka [01...31] yang dapat diartikan sebagai

urutan angka tanggal '01', '02', '03', ..., '29', '30', '31'.

Contoh penulisan benar:

	04
	22

Contoh penulisan salah:

	022
	32

### Bulan

Penulisan tipe data Bulan terbatas pada karakter 2-angka [01...12] yang dapat diartikan sebagai pemetaan nama-nama bulan:

- '01' - 'Januari',
- '02' - 'Februari',
- '03' - 'Maret',
- '04' - 'April',
- '05' - 'Mei',
- '06' - 'Juni',
- '07' - 'Juli',
- '08' - 'Agustus',
- '09' - 'September',
- '10' - 'Oktober',
- '11' - 'November',
- '12' - 'Desember'.

Contoh penulisan benar:

	02
	10

Untuk menghindari permasalahan penggunaan ragam bahasa dan penerjemahannya, dokumen ini tidak merekomendasikan penggunaan label penamaan bulan. Oleh karenanya penulisan nama bulan seperti "Januari", "Februari", "Maret", dan variasinya harus dihindari.

Contoh penulisan salah:

	Februari
	Feb

## Tahun

Penulisan tipe data Tahun terbatas pada karakter 4-angka [0000...9999].

Tanda minus ('-') dapat digunakan sebagai penanda tahun sebelum era umum (SM - Tahun Sebelum Masehi). Penulisan angka tahun yang dipersingkat tidak diperbolehkan.

Contoh penulisan benar:

	-0001
	0354
	1999

*Catatan: Contoh pertama merujuk tahun 1 SM*

Contoh penulisan salah:

	354
	'99

## Tanggal Kalender

Penulisan tipe data Tanggal Kalender menggunakan komponen-komponen karakter dari penulisan informasi Tanggal, Bulan dan Tahun. Secara spesifik penulisan baku tipe data Tanggal mengikuti format sebagai berikut:

[Tahun]-[Bulan]-[Tanggal]

atau

[Tahun]-[Bulan]

Perhatikan karakter tanda pisah ('-') dipakai untuk memisahkan tahun, bulan dan tanggal.

Contoh penulisan benar:

	2014-01-01
	1999-12-31
	1999-12

Contoh penulisan salah:

2014-1-1
01-01-2014
31 Desember 1999

## Waktu

Penulisan tipe data Waktu mengikuti aturan sistem waktu 24-jam. Komponen-komponen penulisannya terdiri dari penunjuk jam, menit dan detik. Khusus penunjuk detik, pencantuman informasi ini bersifat opsional. Secara spesifik penulisan baku tipe data Jam mengikuti format sebagai berikut:

[jam]:[menit]:[detik]

atau

[jam]:[menit]

Perhatikan karakter tanda titik-dua (':') dipakai untuk memisahkan jam, menit dan detik.

Dokumen ini merekomendasikan penulisan waktu yang disertai keterangan zona waktu di Indonesia. Keterangan zona waktu disesuaikan dengan aturan *Coordinated Universal Time* (UTC) yang membagi 3 zona waktu di Indonesia, antara lain:

- Waktu Indonesia Barat (WIB):

[jam]:[menit]:[detik]**+0700**

- Waktu Indonesia Tengah (WITA):

[jam]:[menit]:[detik]**+0800**

- Waktu Indonesia Timur (WIT):

[jam]:[menit]:[detik]**+0900**

Apabila keterangan zona waktu tidak diberikan maka terdapat potensi penulisan informasi waktu yang ambigu. Kesepakatan umum yang sering dijadikan alternatif adalah penerimaan konteks waktu sesuai asal wilayah cakupan data.

Contoh penulisan benar:

09:15:10+0700
---------------

10:15:10+0800
11:15:10+0900
19:15:10
19:15

Contoh penulisan salah:

09.15.10
09.15
9:15
09:15 WIB

## Lokasi

Tipe data Lokasi terbagi menjadi 3 sub-tipe antara lain:

- garis lintang,
- garis bujur,
- alamat

### Garis Lintang

Penulisan tipe data Garis Lintang terbatas pada penggunaan derajat bumi +90 sampai -90. Dokumen ini merekomendasikan tingkat ketelitian minimum 4-angka di belakang koma. Penulisannya menggunakan format angka desimal.

Contoh penulisan benar:

30.0000
-15.6821828

### Garis bujur

Penulisan tipe data Garis Lintang terbatas pada penggunaan derajat bumi +180 sampai -180. Dokumen ini merekomendasikan tingkat ketelitian minimum 4-angka di belakang koma. Penulisannya menggunakan format angka desimal.

Contoh penulisan benar:

110.0000
-62.6821828

## Alamat

Penyusunan nama alamat memerlukan 5 komponen wajib penunjuk lokasi antara lain:

- nama jalan,
- nomor bangunan,
- nama kelurahan atau kecamatan,
- nama kota,
- kode pos.

Adapun komponen lain yang bersifat opsional antara lain:

- nomor blok,
- nomor RT,
- nomor RW,
- nama kompleks,
- provinsi

Secara umum penulisan baku tipe data Alamat mengikuti format sebagai berikut:

[nama kompleks], [nama jalan] No. [nomor], [kelurahan / desa], [kecamatan], [kota] [kode pos]

Kata “Nomor” atau akronim “No.” dipakai untuk memisahkan antara nama jalan dan nomor bangunan. Perhatikan bahwa tanpa kehadiran kata atau akronim pemisah tersebut, penulisan angka selanjutnya akan diasumsikan sebagai bagian dari nama jalan ketimbang nomor bangunan. Karakter spasi tunggal dipakai untuk memisahkan nama kota dan kode pos.

Contoh penulisan benar:

Jl. Pati No. 1, Menteng, Jakarta Pusat 10310
Kompleks Giri Indah, Jl. Dukuh 3 No. 3, Cibodas, Tangerang 15138
Jl. Dukuh 3 No. 3 RT02/RW11, Cibodas, Tangerang 15114, Banten

*Catatan: Contoh ketiga bersifat pilihan karena keterangan RT, RW dan provinsi tidak mutlak dicantumkan.*

Contoh penulisan salah:

Jl. Pati No. 1, Menteng
Jl. Dukuh 3, Cibodas, Tangerang 15138

*Catatan:*

- Contoh pertama salah karena tidak mencantumkan nama kota dan kode pos.
- Contoh kedua salah karena tidak ada nomor rumah. Perhatikan bahwa “Jl. Dukuh 3” tidak

sama dengan "Jl. Dukuh No. 3".

## Telepon

Penyusunan nomor telepon memerlukan 2 komponen wajib antara lain:

- Kode sambungan:
  - Kode area (untuk PTSN),
  - Kode operator (untuk seluler),
- Nomor sambungan.

Adapun komponen lain yang bersifat optional antara lain:

- Kode negara. Umumnya untuk nomor telepon regional, kode negara tidak perlu dicantumkan secara eksplisit.

Secara spesifik penulisan baku tipe data Telepon mengikuti format sebagai berikut:

- Nomor PTSN:  
[kode area] [nomor sambungan]
- Nomor Seluler:  
[kode operator] [nomor sambungan]

Dokumen ini memberikan kebebasan penggunaan variasi format penulisan nomor telepon asalkan untuk selalu konsisten di setiap penulisan datanya.

Contoh penulisan benar:

021 123456789
0815 123456789
(021) 123456789
0815-1234-56789
0039 21 123456789
+39 815 123456789

*Catatan: Dua contoh terakhir menampilkan penulisan tipe data Telepon menggunakan kode negara. Perhatikan variasi penulisan kode negara dapat berupa '00[kode negara]' dan '+[kode negara]'*

## Email

Penyusunan alamat email memerlukan 2 komponen wajib antara lain:

- nama lokal,
- nama domain.

Secara spesifik penulisan baku tipe data Email mengikuti format sebagai berikut:

[nama lokal]@[nama domain]

Karakter ampersat (`@`) dipakai untuk memisahkan nama lokal dan nama domain.

Contoh penulisan benar:

namalokal@example.com
nama.lokal@example.com

Contoh penulisan salah:

namalokalsaja
nama.lokal.example.com
nama.lokal(at)example.com
nama(dot)lokal(at)example(dot)com

## URL

Penyusunan alamat situs URL memerlukan 3 komponen sebagai berikut:

- protokol,
- nama host,
- penunjuk lokasi

Protokol yang umum dipakai untuk alamat situs adalah HTTP dan HTTPS. Secara spesifik penulisan baku tipe data URL mengikuti format sebagai berikut:

http://[nama host]/[penunjuk lokasi]

https://[nama host]/[penunjuk lokasi]

Contoh penulisan benar:

http://example.com
https://example.com/data/01/2013/dataset.csv

Contoh penulisan salah:

example.com
/example.com/data/01/2013/dataset.csv

## Rangkuman Tipe Data

Tabel di bawah ini menampilkan rangkuman contoh-contoh penulisan data secara baku sesuai dengan tipe datanya masing-masing.

Tipe Data	Contoh Penulisan Baku
Teks	"Jakarta Pusat" "SND129:21/2014" "Rp 20.000,00" "Daerah Khusus Ibukota Jakarta (DKI Jakarta, Jakarta Raya) adalah ibu kota negara Indonesia."
Angka	"125" "100000" "-100000" "123.45" "0.000012345"
Logika Binari	"1" "0" "ya" "tidak"
Tanggal	"04" "22"
Bulan	"02" "10"
Tahun	"-0001" "0354" "1999"
Tanggal Kalender	"2014-01-01" "1999-12-31"
Waktu	"09:15:10+0700" "10:15:10+0800" "11:15:10+0900" "19:15:10" "19:15"
Garis Lintang	"30.0000" "-15.6821828"
Garis Bujur	"110.0000" "-62.6821828"

Alamat	"Jl. Pati No. 1, Menteng, Jakarta Pusat 10310" "Kompleks Giri Indah, Jl. Dukuh 3 No. 3, Cibodas, Tangerang 15138"
Telepon	"021 123456789" "0815 123456789" "(021) 123456789" "0815-1234-56789" "+39 21 123456789"
Email	"namaemail@example.com" "nama.email@example.com"
URL	"http://example.com" "https://example.com/data/01/2013/dataset.csv"