

29 Juli 2026

Sumber Data dan Informasi Ilmiah

Pelatihan Teknis Substantif Jabfung ADI Batch 4

Disampaikan oleh: Rulina Rachmawati

Pustakawan Ahli Muda (Direktorat Repositori, Multimedia, dan Penerbitan Ilmiah-BRIN)

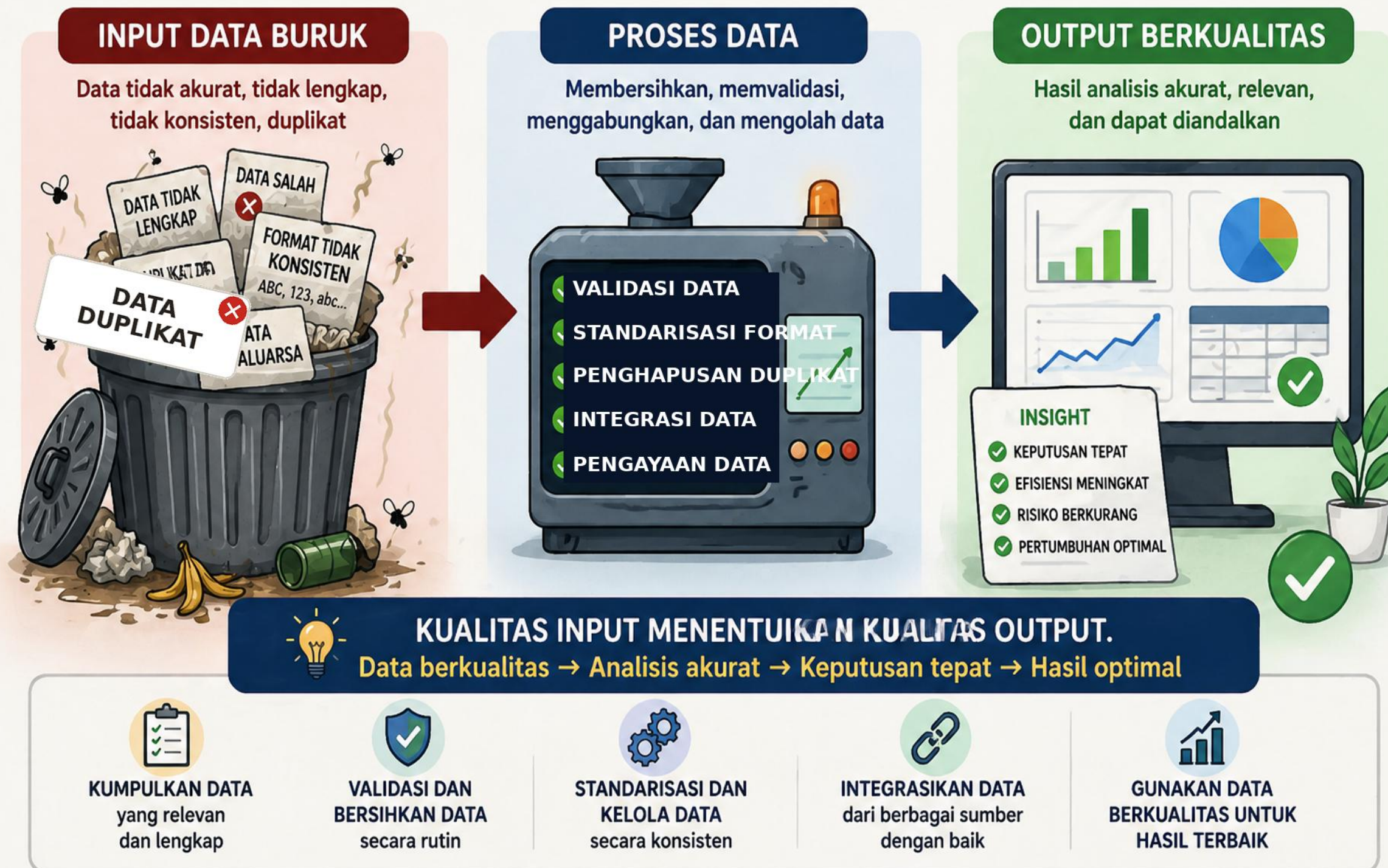
Outline

- 1) Latar Belakang dan Definisi
- 2) Jenis Data dan Informasi Ilmiah
- 3) Sumber Data dan Informasi Ilmiah
- 4) Metadata
- 5) Penelusuran Data dan Informasi Ilmiah
- 6) Evaluasi/ Kriteria Pemilihan Data dan Informasi Ilmiah

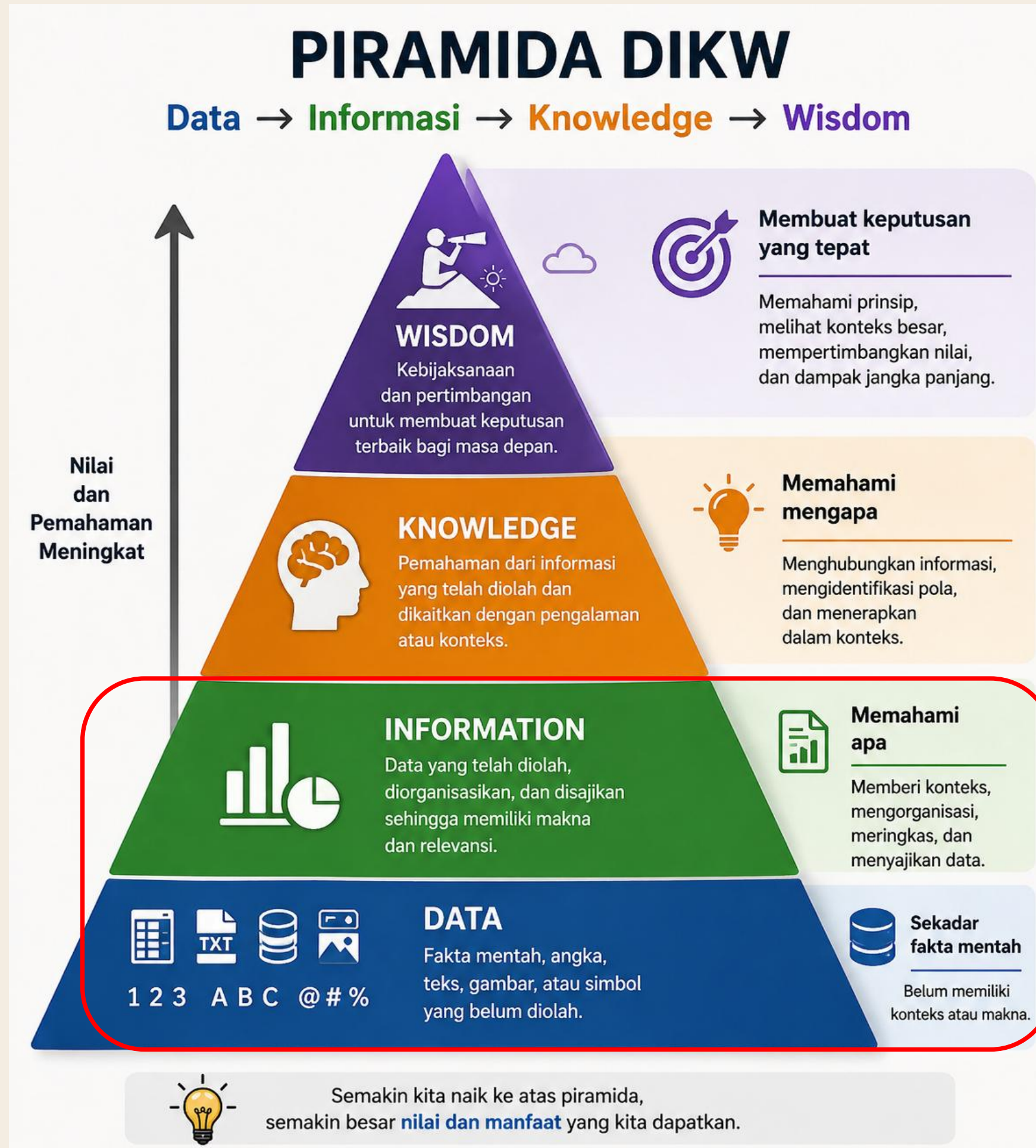
Latar Belakang

GARBAGE IN, GARBAGE OUT

Kualitas analisis data bergantung pada kualitas data yang digunakan.



Definisi

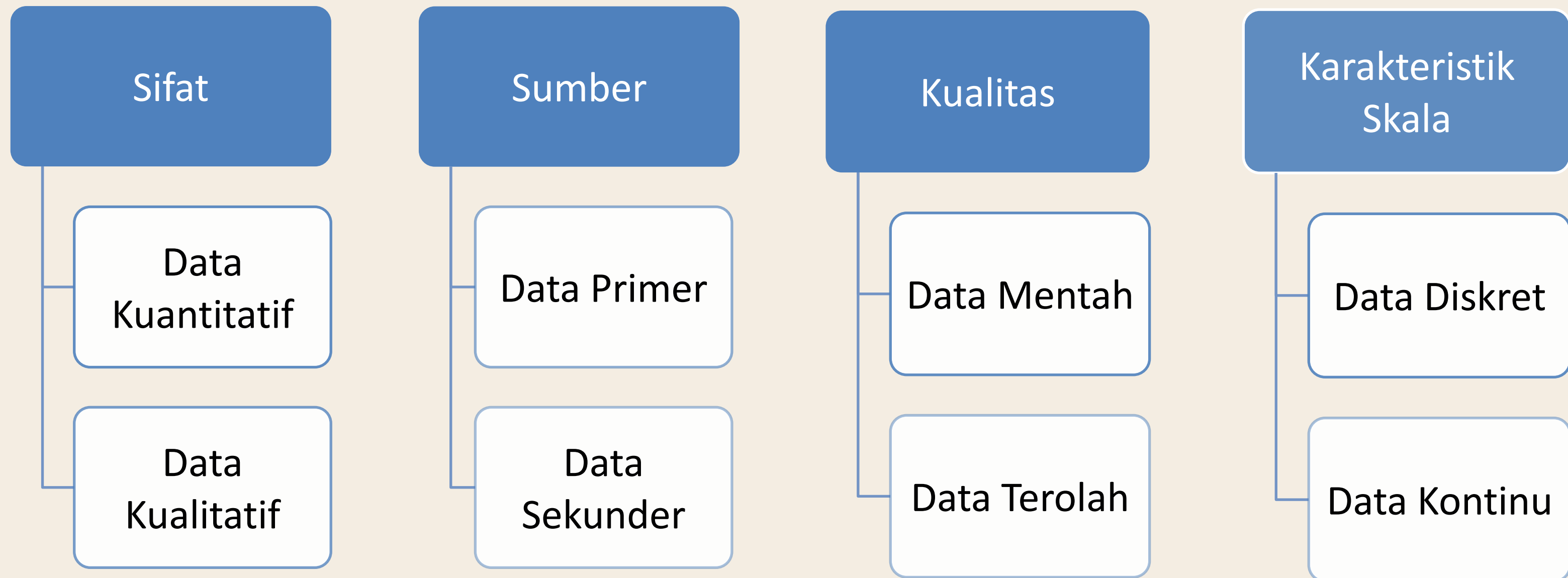


FRAMEWORK DIKW (Data – Information – Knowledge – Wisdom)



Tujuan Akhir: Menghasilkan keputusan yang bernilai, berdampak positif, dan berkelanjutan.

Jenis Data dan Informasi



Sifat

Aspek	Kuantitatif	Kualitatif
Fokus	Angka, statistik	Kata, deskripsi
Metode Pengumpulan	Survei, eksperimen	Wawancara, observasi
Analisis	Statistik, matematis	Identifikasi tema, pola
Tujuan	Pengujian hipotesis, generalisasi	Pemahaman mendalam

Sumber

Aspek	Data Primer	Data Sekunder
Definisi	Data yang dikumpulkan langsung oleh peneliti untuk tujuan spesifik penelitian. 1 2	Data yang sudah ada sebelumnya, dikumpulkan untuk tujuan lain, tetapi digunakan kembali untuk penelitian baru. 1 2 3
Sumber	Diperoleh langsung dari responden atau objek penelitian melalui wawancara, survei, eksperimen, atau observasi. 2 4 5	Diperoleh dari dokumen, publikasi, database, atau catatan yang sudah ada. 2 3 6
Keuntungan	- Data lebih relevan dan spesifik untuk tujuan penelitian. 1 2 - Peneliti memiliki kontrol penuh atas proses pengumpulan data. 7	- Lebih hemat waktu dan biaya karena data sudah tersedia. 1 3 8 - Dapat mencakup data dalam skala besar atau historis. 6 8
Kekurangan	- Proses pengumpulan data memakan waktu dan biaya tinggi. 3 7 - Rentan terhadap bias jika tidak direncanakan dengan baik. 7	- Mungkin tidak sepenuhnya relevan dengan tujuan penelitian. 6 9 - Kualitas data tergantung pada metode pengumpulan awal. 7 9
Contoh	- Hasil wawancara langsung dengan responden. 5 - Data eksperimen yang dikumpulkan di laboratorium. 4	- Data sensus, laporan pemerintah, atau catatan medis elektronik. 1 6 10

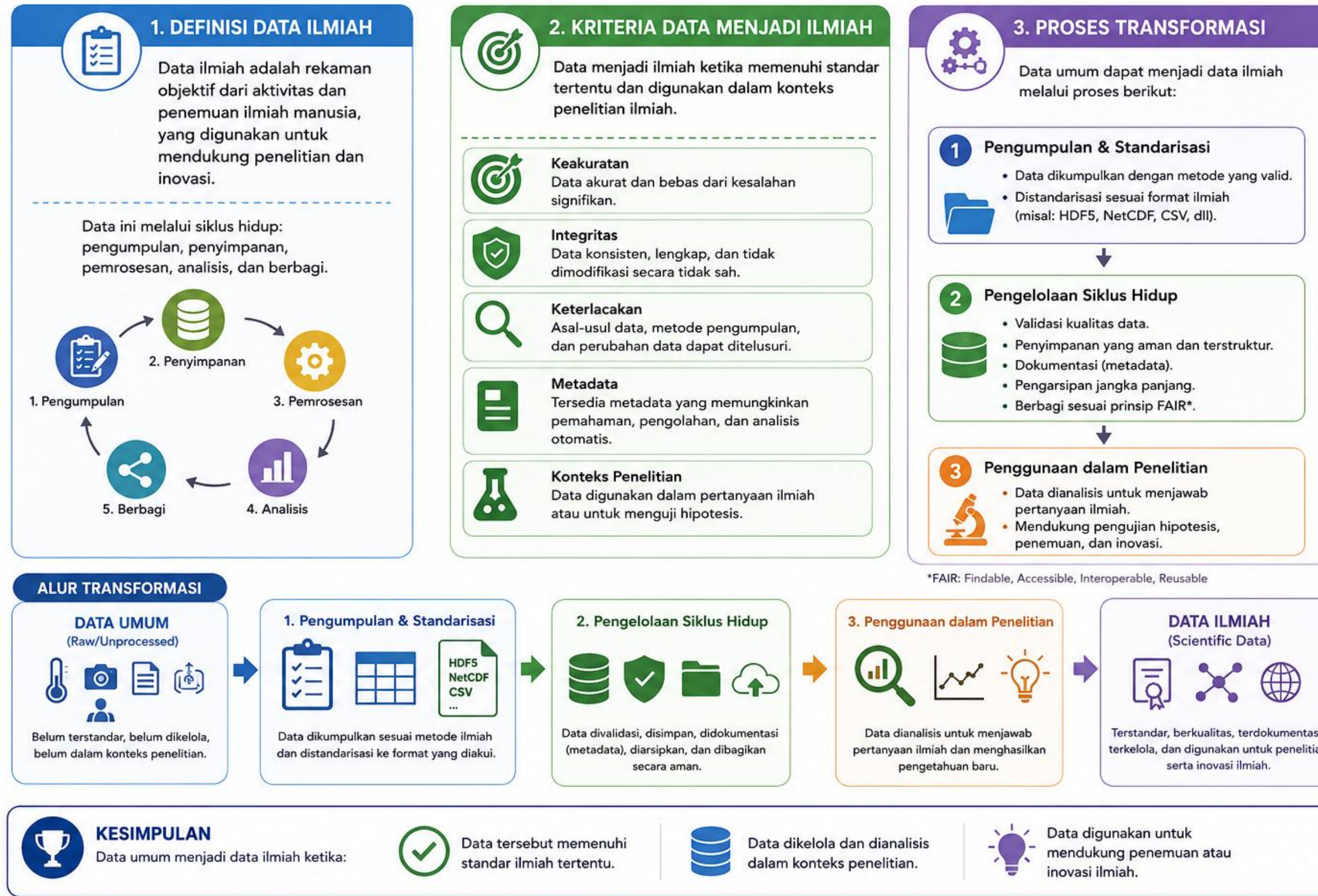
Kualitas

Aspek	Data Mentah (<i>Raw Data</i>)	Data Terolah (<i>Processed Data</i>)
Pengertian	Data yang baru dikumpulkan dari sumbernya dan belum dianalisis atau diinterpretasikan.	Data yang telah dibersihkan, diolah, dianalisis, dan disajikan sehingga memiliki makna.
Bentuk	Angka, teks, rekaman, hasil observasi, atau hasil pengukuran asli.	Tabel, grafik, statistik, ringkasan, temuan penelitian, atau laporan.
Tujuan	Sebagai bahan dasar analisis.	Mendukung pengambilan keputusan dan menghasilkan informasi.
Tingkat Keterbacaan	Sulit dipahami secara langsung.	Lebih mudah dipahami dan diinterpretasikan.
Proses yang Diperlukan	Belum melalui validasi, pembersihan, atau analisis.	Sudah melalui pembersihan data, pengelompokan, analisis, dan visualisasi.
Nilai Informasi	Rendah karena belum memberikan makna yang jelas.	Tinggi karena telah menghasilkan informasi atau pengetahuan.

Karakteristik Skala

Aspek	Data Diskret	Data Kontinu
Pengertian	Data yang memiliki nilai pasti dan tidak terdapat nilai di antara dua nilai yang berurutan. Umumnya diperoleh dari hasil menghitung (counting).	Data yang bersifat berkelanjutan dan memiliki kemungkinan nilai di antara dua nilai yang berurutan. Umumnya diperoleh dari hasil pengukuran (measuring).
Karakteristik Nilai	Nilainya terpisah dan dapat dihitung satu per satu.	Nilainya berada dalam suatu rentang dan dapat terus dibagi menjadi nilai yang lebih kecil.
Pecahan/Desimal	Biasanya berupa bilangan bulat dan tidak memiliki nilai pecahan.	Dapat berupa bilangan pecahan atau desimal.
Metode Perolehan	Hasil penghitungan jumlah objek, individu, atau kejadian.	Hasil pengukuran menggunakan alat ukur tertentu.
Jumlah Kemungkinan Nilai	Terbatas atau dapat dihitung.	Tidak terbatas dalam suatu interval tertentu.
Contoh Umum	Jenis kelamin, jumlah orang, jumlah kendaraan, jumlah buku, jumlah artikel, jumlah kunjungan perpustakaan.	Berat badan, mata uang, panjang, suhu, waktu, jarak, tinggi badan.

TRANSFORMASI DATA UMUM MENJADI DATA ILMIAH



Sumber Data dan Informasi Ilmiah



Instansi Pemerintah

LKPP; Satu Data Indonesia, Kementerian PPN/Bappenas; BNPB; BIG; BPS; Pemerintah USA.



Repositori Data

RIN; Kaggle; re3data; World Bank.



Sumber Internet

Google; Datahub; Dimensions; Open Contracting Data Registry.



Dokumen & Arsip

Literatur abu-abu (*grey literature*), laporan internal, arsip nasional.



Industri / Swasta

Data Tempo; Databoks; Website perusahaan masing-masing.



Eksperimen

Simulasi, observasi langsung, dan data laboratorium.

Sumber data	Situs
Instansi Pemerintah	LKPP: https://data.lkpp.go.id/dataset/ Satu Data Indonesia, Kementerian PPN/Bapennas: https://data.go.id/dataset BNPB: https://dibi.bnpb.go.id/superset/dashboard/1/?standalone=0&expand_filters=0 BIG: https://geoportal.big.go.id/#/katalog BPS: https://www.bps.go.id/ Pemerintah USA: https://data.gov/
Repositori Data	Figshare: https://figshare.com/search?q=dataset Mendeley data: https://data.mendeley.com/ Harvard dataverse: https://dataverse.harvard.edu/ RIN: https://rin.brin.go.id/ Kaggle: https://www.kaggle.com/datasets re3data: https://www.re3data.org/search World Bank: https://datacatalog.worldbank.org/search?sort=last_updated_date%20desc
Sumber Internet	Google: https://datasetsearch.research.google.com/ Datahub: https://datahub.io/collections Dimensions: https://www.dimensions.ai/ Open Contracting Data Registry: https://data.open-contracting.org/en/search/
Industri/ Swasta	Data Tempo: https://data.tempo.co/front/datacenter/ Databoks: https://databoks.katadata.co.id/ Website perusahaan masing-masing

ELECTRONIC RESOURCES

Database



NATURE

IEEE Xplore®



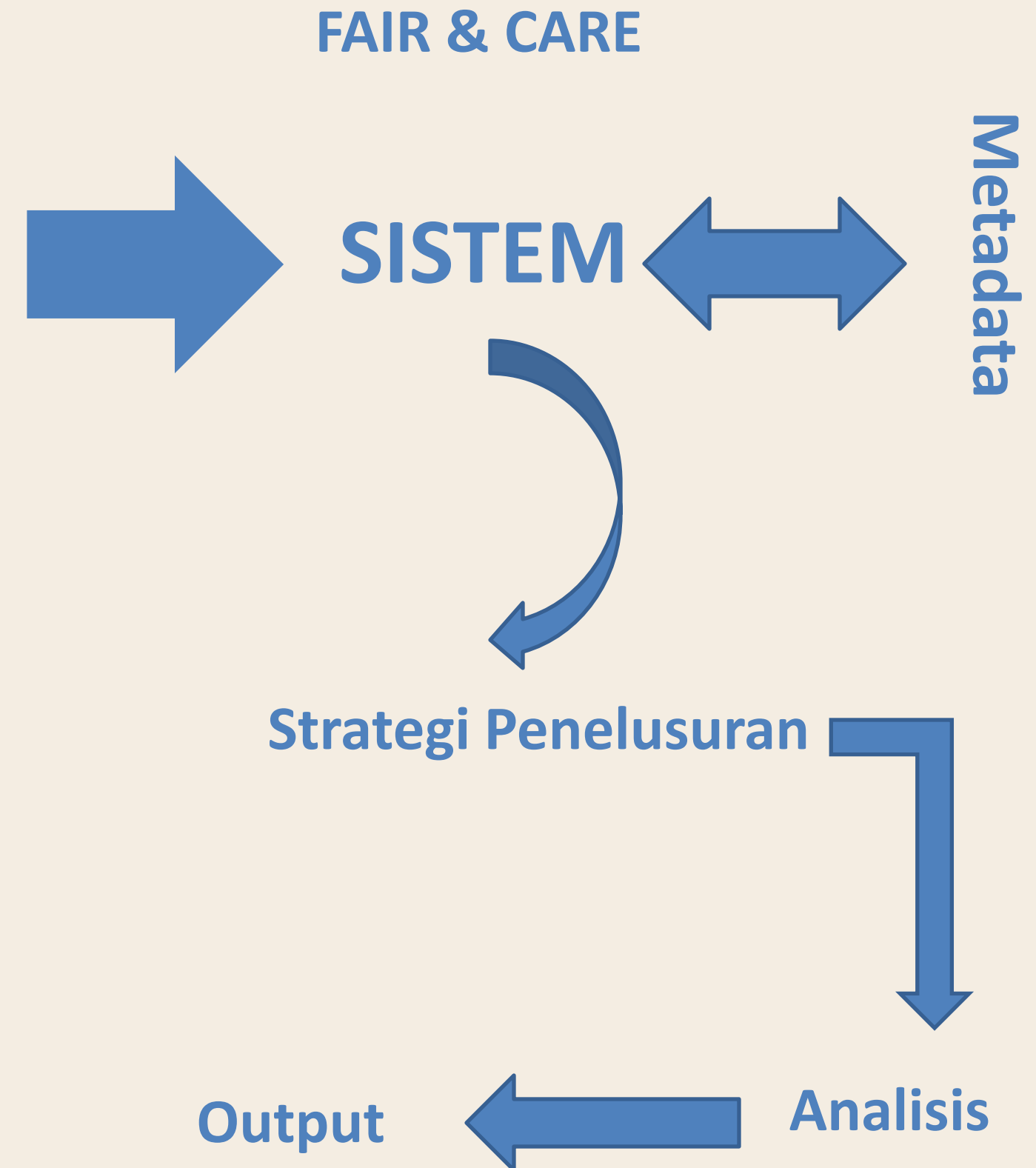
IOP Publishing



Indexer



Alur Pemanfaatan Data dan Informasi Ilmiah



FAIR

Prinsip pengelolaan data agar mudah ditemukan, diakses, dioperasikan, dan digunakan kembali.



1. Findable (Dapat Ditemukan)

- Metadata standar
- Identifikasi unik
- Indeksasi



2. Accessible (Dapat Diakses)

- Open access
- Kebijakan data sharing



3. Interoperable (Dapat Dioperasikan)

- Penggunaan format standar



4. Reusable (Dapat Digunakan Kembali)

- Lisensi
- Kualitas metadata

CARE

Prinsip etika dan tata kelola data yang menghormati hak, kepentingan, dan kontrol komunitas.



1. Collective Benefit

- Tujuan sosial, keberlanjutan komunitas adat



2. Authority to Control

- Kedaulatan data, hak pengelolaan berbasis persetujuan komunitas



3. Responsibility

- Akuntabilitas pengguna data, transparansi proses pengelolaan data



4. Ethics

- Penghormatan terhadap hak adat, keamanan dan privasi data

FAIR menekankan kualitas dan keterbukaan data, sedangkan **CARE** menekankan etika, hak komunitas, dan tanggung jawab penggunaan data.

Metadata

Data yang menjelaskan, mendeskripsikan, mengidentifikasi, dan memberikan informasi mengenai data lainnya sehingga data tersebut dapat ditemukan, dipahami, dikelola, diakses, dan digunakan kembali secara efektif

- Kunci dalam pengelolaan data digital yang semakin kompleks
- Dengan standar yang baik, metadata dapat meningkatkan efisiensi pencarian, pengelolaan, dan penggunaan data
- Metadata juga berperan penting dalam mendukung integrasi data lintas sistem dan domain

Contoh Metadata
Dublin Core

Judul artikel: Metadata: A New Word for an Old Concept
Penulis: A. Yousefi dan S. Yousefi
Tahun: 2007
Jurnal: Library Philosophy and Practice

Metadata Dublin Core

Elemen Dublin
Core

Isi Metadata

Title	Metadata: A New Word for an Old Concept
Creator	Yousefi, A.; Yousefi, S.
Subject	Metadata; bibliographic description; library science; information organization
Description	Artikel ini membahas konsep metadata sebagai istilah baru untuk konsep lama dalam pengelolaan dan deskripsi informasi bibliografis.
Publisher	Library Philosophy and Practice
Contributor	—
Date	2007
Type	Text; Journal Article
Format	Electronic article
Identifier	—
Source	Library Philosophy and Practice
Language	English
Relation	—
Coverage	—
Rights	Hak cipta mengikuti ketentuan penerbit atau jurnal

Contoh Metadata di RIN

Citation Metadata ^	
Persistent Identifier ?	hdl:20.500.12690/RIN/PWW2IU
Publication Date ?	2026-03-11
Title ?	LOKAMATH-QA: Dataset Penalaran Soal Matematika Berbasis Bahasa Lokal Indonesia
Alternative Title ?	LOKAMATH-QA: A Multilingual Mathematical Reasoning Dataset for Indonesian Local Languages
Author ?	Nimah, Ifitahu (Pusat Riset Sains Data dan Informasi BRIN) - ORCID: 0000-0003-2350-0481 Yuyun (Pusat Riset Sains Data dan Informasi BRIN) - ORCID: 0000-0003-4936-1862 Kokoy Siti Komariah (Pusat Riset Sains Data dan Informasi BRIN) - ORCID: 0000-0003-2618-3039 Aria Bisri (Pusat Riset Sains Data dan Informasi BRIN) - ORCID: 0009-0002-0650-5520 Nurfaedah Jufri (Universitas Handayani Makassar) - ORCID: 0009-0003-3218-2072
Point of Contact ?	Use email button above to contact. Nimah, Ifitahu (Pusat Riset Sains Data dan Informasi BRIN)
Description ?	Deskripsi dalam Bahasa Indonesia: Dataset terdiri dari 500 pasangan pertanyaan, penalaran, dan jawaban dari soal matematika multi bahasa, yaitu Bahasa Indonesia, Jawa, Sunda, dan Bugis. Dataset ini merupakan pengembangan dari dataset penalaran soal matematika dalam Bahasa Inggris – MATH500 (Hendrycks et al., 2021), yang diterjemahkan dengan menggunakan OpenAI GPT-4.1 dan diperbaiki lebih lanjut oleh penutur bahasa asli. Dataset ini dikembangkan untuk menjembatani riset mengenai analisis dan evaluasi model bahasa besar multibahasa (multilingual) untuk domain pendidikan atau penalaran soal matematika, terutama untuk bahasa-bahasa bersumber daya rendah (low resource language) yang mencakup tiga (3) bahasa lokal di Indonesia: Bahasa Jawa, Sunda, dan Bugis. (2026-03-09) English description: Dataset is composed of 500 question answering pairs for each target language, including the reasoning process of open source Large Language Models (LLMs) in Indonesian, Javanese, Sundanese, and Buginese language. This dataset was constructed from 500 English samples of math reasoning task Q&A -- Math500 (Hendrycks et al., 2021), by utilizing OpenAI GPT-4.1 for automatic translation from Indonesian language to target languages and further refined by native speakers of Javanese, Sundanese, and Buginese language. This dataset was constructed to bridge the research on analyzing and evaluating LLMs for low resource languages, particularly local languages in Indonesia.
Subject ?	Computer and Information Science
Keyword ?	mathematical reasoning (Wikidata) https://www.wikidata.org/wiki/Q122766417 low-resource languages (LCSH) https://id.loc.gov/authorities/subjects/sh2022006704.html large language model (Wikidata) https://www.wikidata.org/wiki/Q115305900 natural language processing (LCSH) https://id.loc.gov/authorities/subjects/sh88002425.html indonesian language (LCSH) https://id.loc.gov/authorities/subjects/sh85065767.html javanese language (LCSH) https://id.loc.gov/authorities/classification/PL5161-PL5179.html sundanese language (LCSH) https://id.loc.gov/authorities/classification/PL5451-PL5454.html bugis language (LCSH) https://id.loc.gov/authorities/classification/PL5271-PL5271.95.html
Related Publication ?	Xiao Xiao, Ifitahu Ni'mah, YuyunWabula, Mykola Pechenizkiy, and Meng Fang. MATH-IDN: A multilingual mathematical problem solving dataset featuring local languages in indonesia. In 19th Conference of the European Chapter of the Association for Computational Linguistics, 2026

Contoh Metadata
dari Mendeley data

Data for: Public Procurement Status and Improvement Needs in Saudi Arabia: Local Content and Saudization

Published: 23 June 2026 | Version 2 | DOI: 10.17632/m7rp6743y7.2

Contributor:  Alaa Khodroj

Description

This dataset presents key metrics from Saudi Arabia’s public procurement landscape, including localization achievements under the IKTVA program, Saudization trends, SME participation, and procurement’s share of GDP. It supports the analysis in the paper “Public Procurement Status and Improvement Needs in Saudi Arabia”.

Keywords: Saudi Arabia, Public Procurement, IKTVA, Saudization, SME Inclusion, Etimad, Vision 2030

Download All 26.7 KB



Files

 data.csv	117 B 
 figure.png	30 KB 
 README.txt	677 B 

Categories

Federal Government, Saudi Arabia, Procurement System, Public Procurement

Licence

CC BY 4.0

[Learn more](#)

Latest version

Version 2

Published: 23 Jun 2026

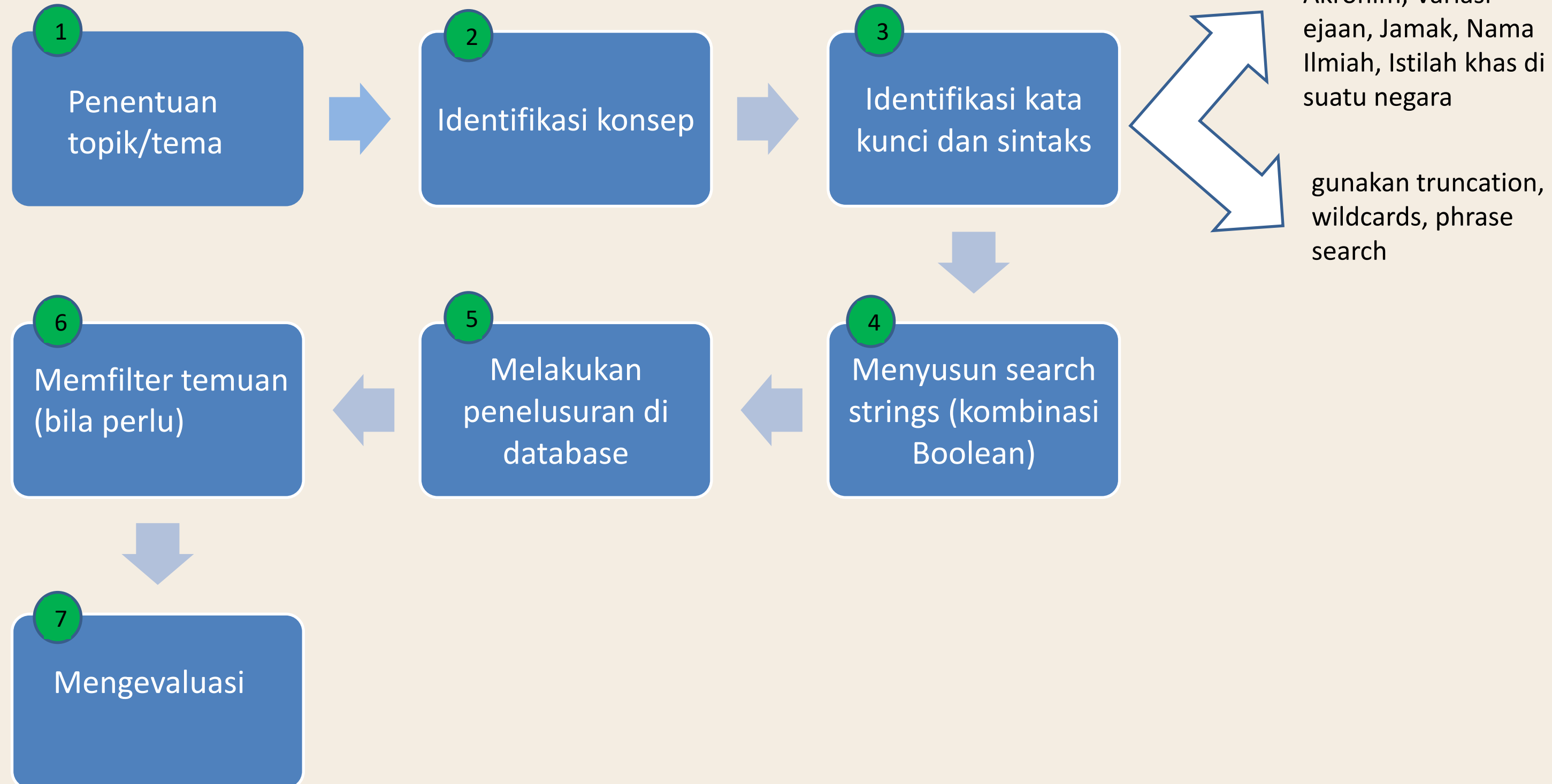
DOI: 10.17632/m7rp6743y7.2

Cite this dataset

Khodroj, Alaa (2026), “Data for: Public Procurement Status and Improvement Needs in Saudi Arabia: Local Content and Saudization”, Mendeley Data, V2, doi: 10.17632/m7rp6743y7.2

 [Copy to clipboard](#)

Strategi Penelusuran Informasi Ilmiah



1 & 2. Penentuan Topik dan Konsep

Topik: Nuclear waste (HLW) management strategies

Konsep: Nuclear waste (HLW) management strategies

Topik: The role of dairy in the diets of children with asthma

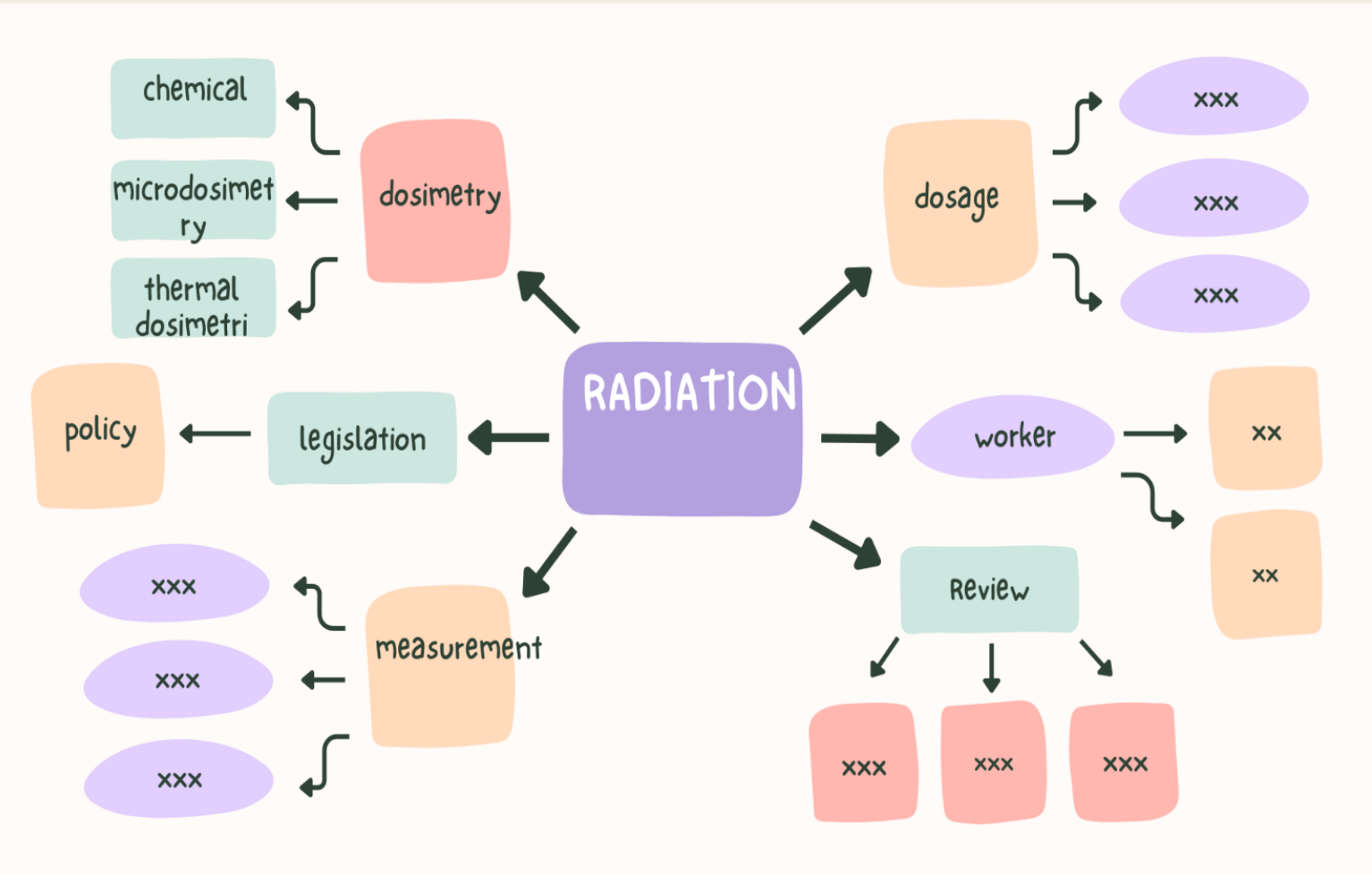
Konsep: the role of dairy in the diets of children with asthma

3. Identifikasi Kata Kunci dan Sintaks

Gunakan :

- Mind mapping
- Thesaurus :
 - Multidisiplin <https://id.loc.gov/search/>
<https://keywordtool.io/>
 - Bidang pendidikan <https://eric.ed.gov/?faq-thesaurus>
 - Bidang kedokteran/ Kesehatan Masyarakat <https://www.ncbi.nlm.nih.gov/mesh/>
 - Bidang nuklir <https://inis.iaea.org/browser/thesaurus>
- Kamus
- Cari satu atau beberapa literatur relevan yang banyak disitir → ambil terms yang umum dipakai ---> author keywords
- ChatGPT

Sinonim	01	curriculum = curricula = syllabus; salafism = salafi salafist = salafis = salafiya = salafiyya
Akronim	02	ADHD = Attention Deficit Hyperactivity Disorder; Diabetes mellitus = DM
Variasi Ejaan	03	organization, organisation;
Jamak	04	syllabi, syllabus, syllabuses
Scientific & Common names	05	green tea = camellia sinensis
Istilah khas di suatu negara	06	dissertation (US); thesis (UK)



Contoh	
Irrelevant	Inappropriate, Unrelated
Observe	Examine, Study, Scrutinize
Radiation dosimetry	BT Nuclear counters Radiation—Measurement
	NT Chemical dosimetry Dosimeters Microdosimetry Photographic dosimetry Thermal dosimetry Thermoluminescence dosimetry
Radiation	RT Radiologi

Truncation, Wildcards, Phrase search

Simbol	Fungsi	Keywords	Kata-kata yang mungkin ditemukan
Truncation *	untuk menemukan variasi kata yang mengandung suku kata yang dipenggal	procur*	procure, procuring, procured, procurement, procurements, dll.
Wildcards # atau ?	digunakan di akhir kata atau di tengah kata untuk menemukan variasi ejaan	col?r	color, colour
		organi?ation	organization, organisation
Exact phrase “ ” atau { }	untuk menemukan frasa yang sama persis urutannya	“3D printing”	"3D printing"
		{3D printing?}	apakah kelebihan 3D printing?

* dan “ ” umum digunakan di banyak database

(?, #, { }) mungkin berbeda antar database, lihat kembali search tips/ search guidance

Contoh hasil temuan menggunakan truncation, wildcards atau phrase search

Terms	Kata-kata yang muncul	Jumlah temuan
procur*	procure, procuring, procured, procurement, procurements, dll.	114,653
procurement	procurement	89,556
wom#n	WOMN	4
wom?n	Women; woman	531,981
“3D printing”	3D printing, 3d-printing	15,779
{3D printing}	3D printing	14,738
{3D printing?}	3D printing?	10

4. Menyusun Search Strings Menggunakan Boolean Operator

OR :

- Menghubungkan kata kunci dengan konsep yang sama
- Memperluas hasil penelusuran
- Menemukan literatur yang mengandung minimal 1 kata kunci yang digunakan
- Contoh: "Government procurement" OR "Public procurement" --> menemukan literatur yang ada salah satu atau kedua kata

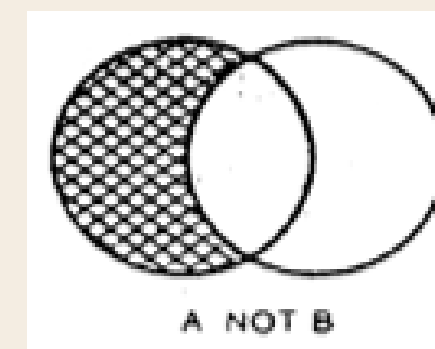
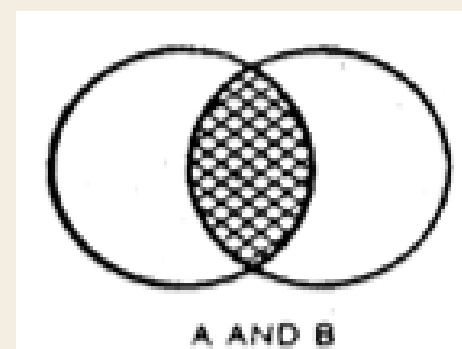
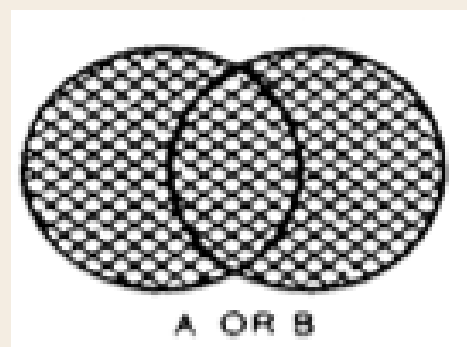
AND :

- Mengkombinasikan konsep yang berbeda
- Mempersempit hasil penelusuran
- Contoh: "Government procurement" AND Corruption --> menemukan hanya literatur tentang korupsi terkait pengadaan barang/jasa pemerintah

NOT :

- tidak menemukan kata setelah NOT
- untuk membatasi penelusuran

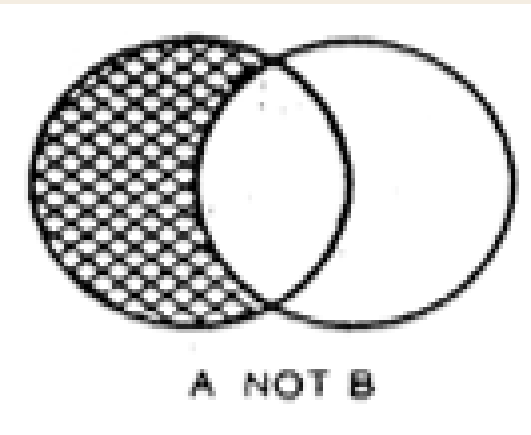
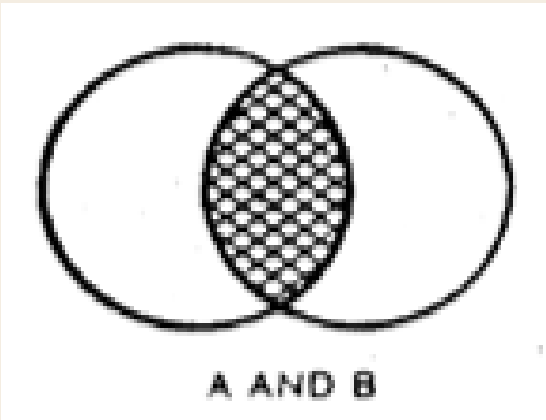
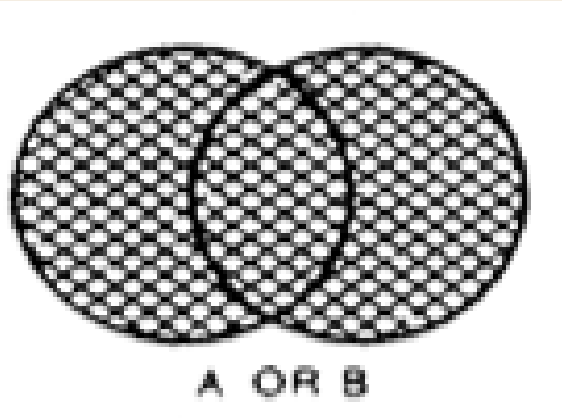
Contoh: Indonesia NOT Australia--> menemukan literatur tentang/ lokasi Indonesia saja.



Contoh Menyusun Search Strings Menggunakan Boolean Operator

Nuclear waste (HLW) management strategies

Konsep	Kata kunci	Search strings
Nuclear waste	Nuclear waste, spent fuel, spent nuclear fuel, radioactive waste	("Nuclear waste" OR "spent fuel" OR "spent nuclear fuel" OR "radioactive waste") AND (hlw OR "High level" OR "high-level" OR "long term" OR "long-range") AND (management OR program OR practice* OR technolog* OR predict* OR forecast*)
HLW	HLW, High level, high-level, long term, long-range	
Management strategies	Management, Program, Practices, Technologies, Prediction, Predicting, Forecasting	



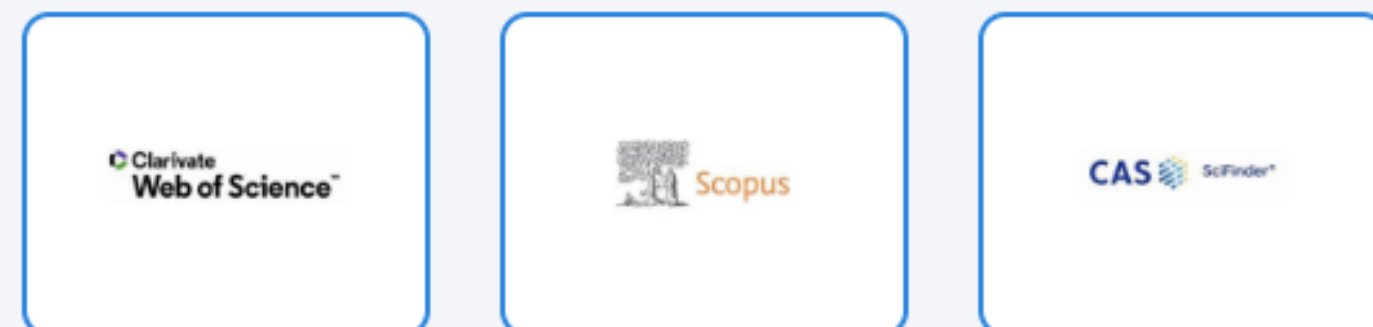
5. Melakukan Penelusuran Literatur

ELECTRONIC RESOURCES

Database



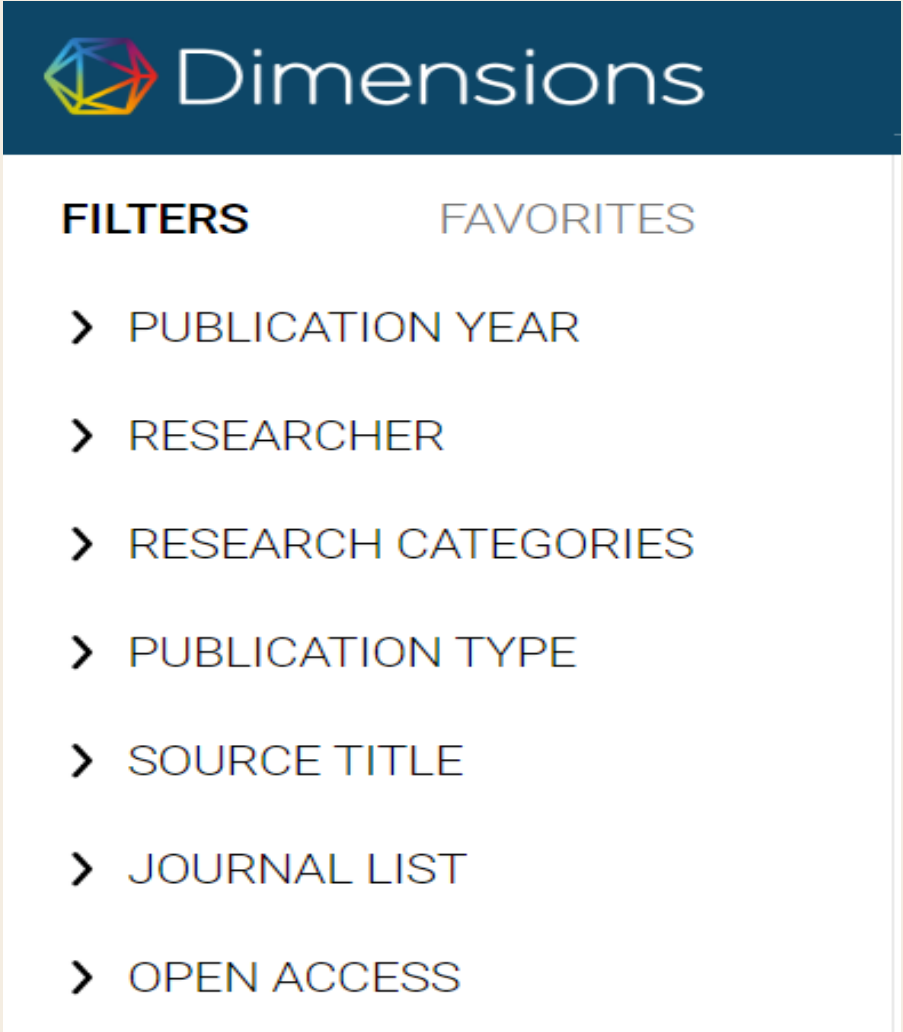
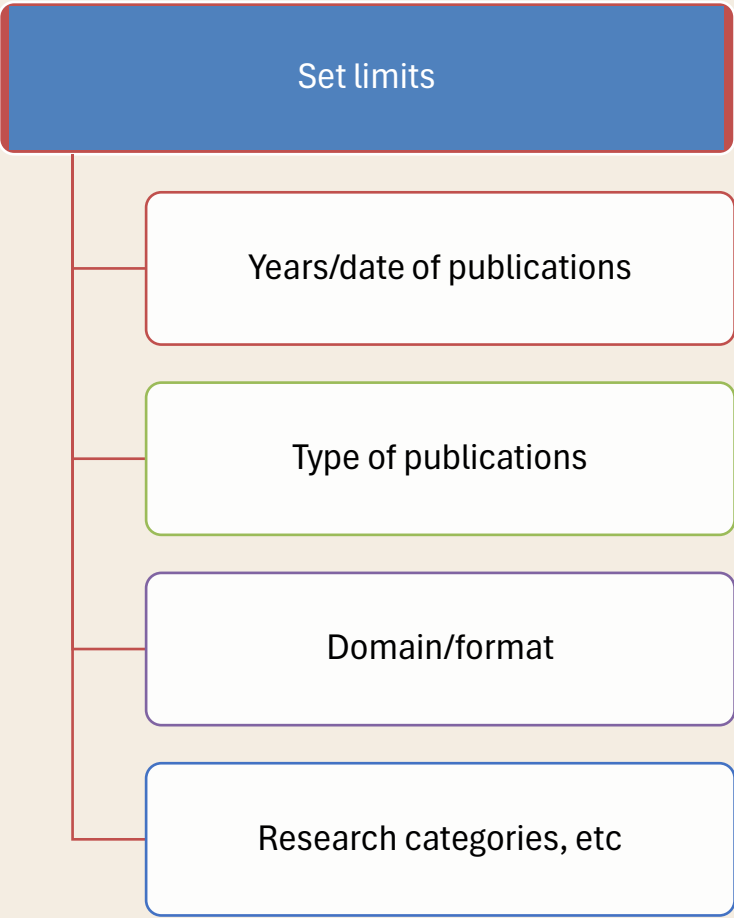
Indexer



Database free access lainnya:

- Google Scholar <https://scholar.google.com/>
- Lens database <https://www.lens.org/>
- Dimensions <https://www.dimensions.ai/> (login untuk bisa akses)

6. Filtering/ Pembatasan (bila diperlukan)



<https://app.dimensions.ai/discover/publication>



<https://pubmed.ncbi.nlm.nih.gov/?term=radiation>

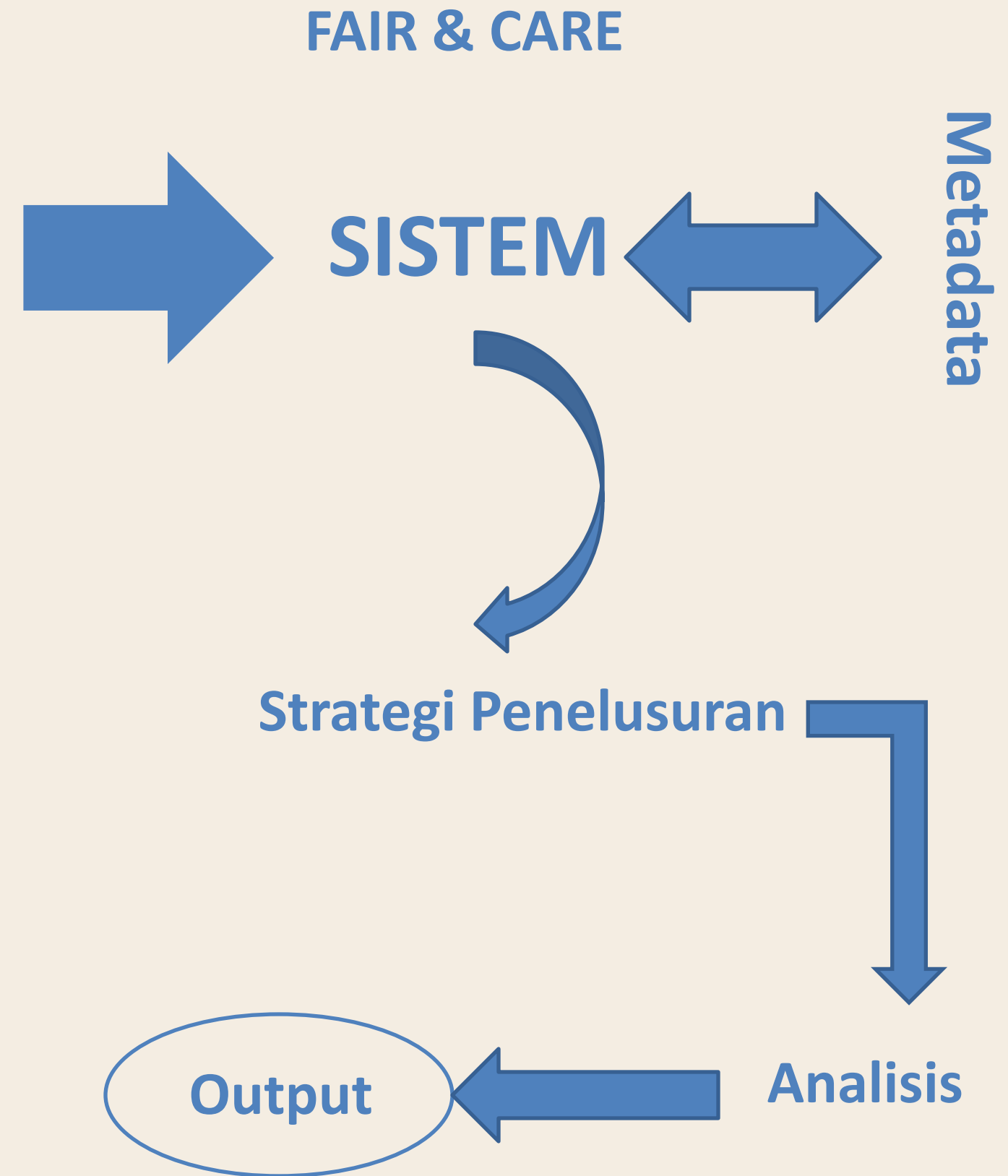
7. Evaluasi Data dan Informasi Ilmiah

- ✓ Kualitas
Akurasi, keandalan, validitas, konsistensi data, Kelengkapan (no missing values)
- ✓ Relevansi
Kesesuaian dengan pertanyaan penelitian/ kebutuhan
- ✓ Aksesibilitas
Kemudahan akses teknis dan biaya
- ✓ Cakupan
Wilayah geografis, waktu, atau populasi yang sesuai
- ✓ Kemutakhiran (tergantung kebutuhan penelitian)
Data terbaru untuk konteks yang dinamis

Alur Pemanfaatan Data dan Informasi Ilmiah



Paper,
Ringkasan Data
Policy brief, dll



Ringkasan Data

Data in Brief 66 (2026) 112684



Contents lists available at ScienceDirect

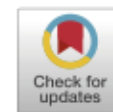
Data in Brief

journal homepage: www.elsevier.com/locate/dib



Data Article

Complete genome dataset of *Flavobacterium* sp. strain PL002 isolated from Antarctic *Porphyra* algae



Jennifer Charles Labo^a, Paris Lavin^b, Hui Yin Fan^c, Mohd Faizal Abu Bakar^d, Nur Athirah Yusof^{a,*}

^a Biotechnology Research Institute, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia

^b Departamento de Biotecnología, Facultad de Ciencias del Mar y Recursos Biológicos, Universidad de Antofagasta, 601 Avenida Angamos, Antofagasta 1270300, Chile

^c Food Security Research Laboratory, Faculty of Food Science and Nutrition, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia

^d Advanced Genomics and Bioinformatics, Malaysia Genome and Vaccine Institute, National Institutes of Biotechnology Malaysia, Jalan Bangi, 43000 Kajang, Selangor, Malaysia

BACA: Jurnal Dokumentasi dan Informasi *Special Issue-Data in Brief for Repositori Ilmiah 2024*

ISSN 0125-9008 (Print); ISSN 2301-8593 (Online)

DOI: 10.55981/j.baca.2024.7785

SK Dirjen Dikti Ristek - Kemdikbudristek No. 105/E/KPT/2022 (Peringkat 2 SINTA)



Dataset VR (*virtual reality*) laboratorium komputer BRIN 2 di Repositori Ilmiah Nasional (RIN)

Arvy Herdianto Firmansyah^{1*}; Abdurrahman Prasetyadi²; Angga Arvianto Hadi³.

^{1,2,3} Badan Riset dan Inovasi Nasional

*Korespondensi: arvyhfirmansyah@gmail.com

Diajukan: 25-Sep-2024; Direview: 21-Okt-2024; Diterima: 19-Nov-2024; Direvisi: 18-Nov-2024

ABSTRACT

The rapid advancement of animation and 3D modeling technology, particularly in virtual reality (VR), has significantly transformed digital interactions. This article discusses the application of VR in computer laboratory simulations for research, experimentation, and data analysis. VR offers interactive 3D visualization, allowing users to view objects and spaces from various perspectives. The research method employed photogrammetry techniques at the BRIN Science and Technology Laboratory in Bandung to gather data. The collected data was then processed using Blender and Unity software, resulting in high-quality 3D models. The VR Lab Computer dataset from BRIN consists of 31 files, including 16 images and 15 3D object models that can be accessed through specific applications. The visualizations produced provide an accurate representation of the laboratory space, equipment, and tools available. The findings of this research indicate the significant potential of VR in education and research, as well as benefits for individuals with physical or geographical limitations. Thus, VR not only enhances the learning experience but also improves the efficiency of repeating experiments.

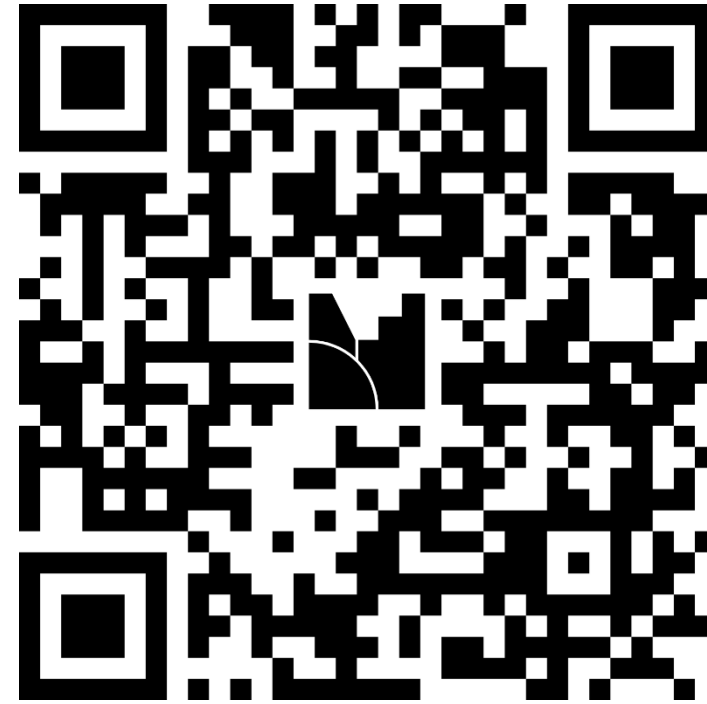
<https://www.sciencedirect.com/journal/data-in-brief/about/aims-and-scope>

<https://ejournal.brin.go.id/baca/issue/view/922>

Terima Kasih

Semoga Bermanfaat 

QUIZ



<https://www.menti.com/al17aiaytdup>