

SUMBER DATA DAN INFORMASI ILMIAH

NOORIKA RETNO WIDURI

Disampaikan pada Pelatihan Teknis Substantif Jabfung ADI Batch 1, 6 Mei 2026

OUTLINE

1. Latar Belakang
2. Jenis dan Sumber data Ilmiah
3. Prinsip Data
4. Metadata
5. Penelusuran data dan informasi ilmiah
6. Kriteria pemilihan data dan informasi ilmiah



BRIN
BADAN RISET
DAN INOVASI NASIONAL

LATAR BELAKANG GARBAGE IN, GARBAGE OUT

Kualitas analisis data bergantung pada kualitas data yang digunakan.

INPUT DATA BURUK

Data tidak akurat, tidak lengkap, tidak konsisten, duplikat



PROSES DATA

Membersihkan, memvalidasi, menggabungkan, dan mengolah data



OUTPUT BERKUALITAS

Hasil analisis akurat, relevan, dan dapat diandalkan



KUALITAS INPUT MENENTUKAN KUALITAS OUTPUT.

Data berkualitas → Analisis akurat → Keputusan tepat → Hasil optimal



KUMPULKAN DATA
yang relevan
dan lengkap



**VALIDASI DAN
BERSIHKAN DATA**
secara rutin



**STANDARISASI DAN
KELOLA DATA**
secara konsisten



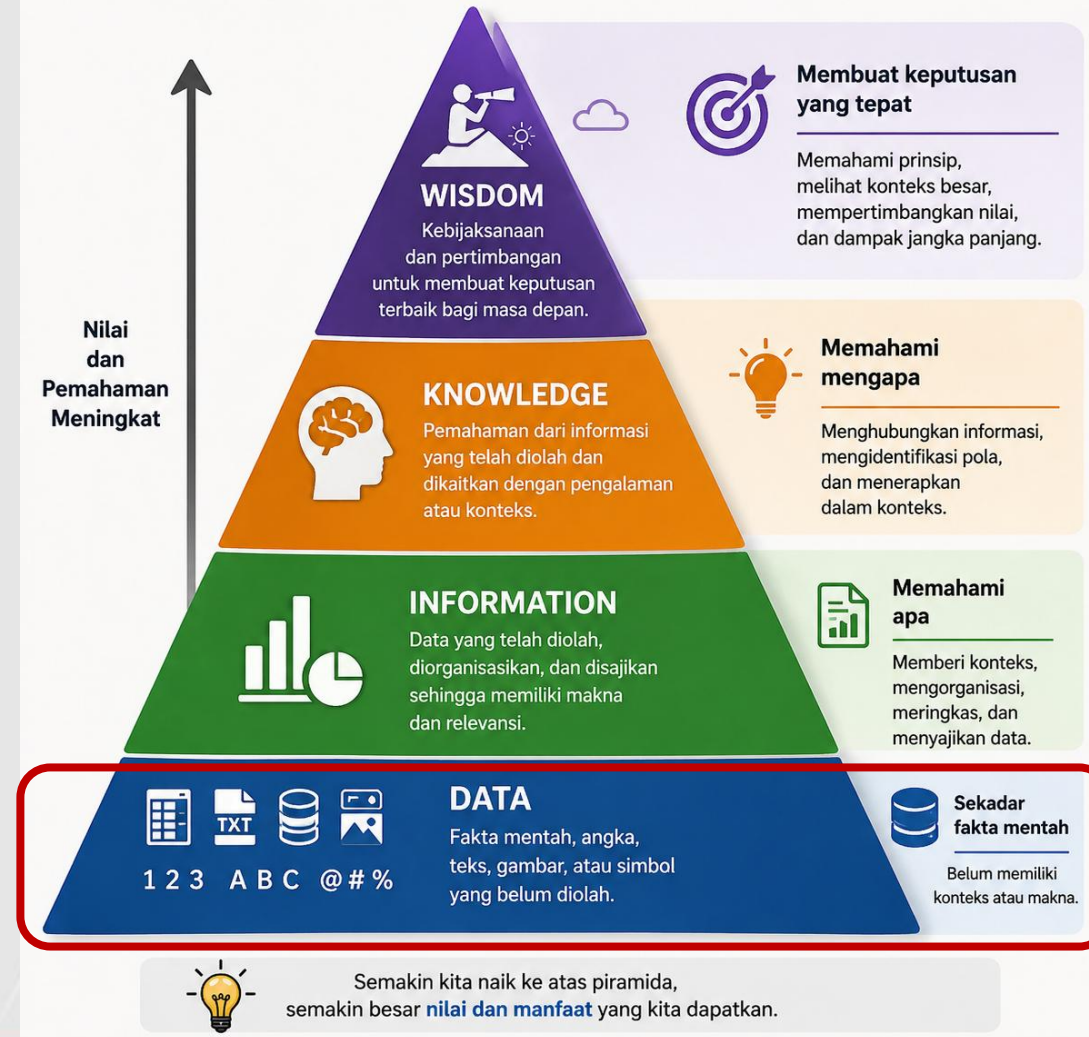
INTEGRASIKAN DATA
dari berbagai sumber
dengan baik



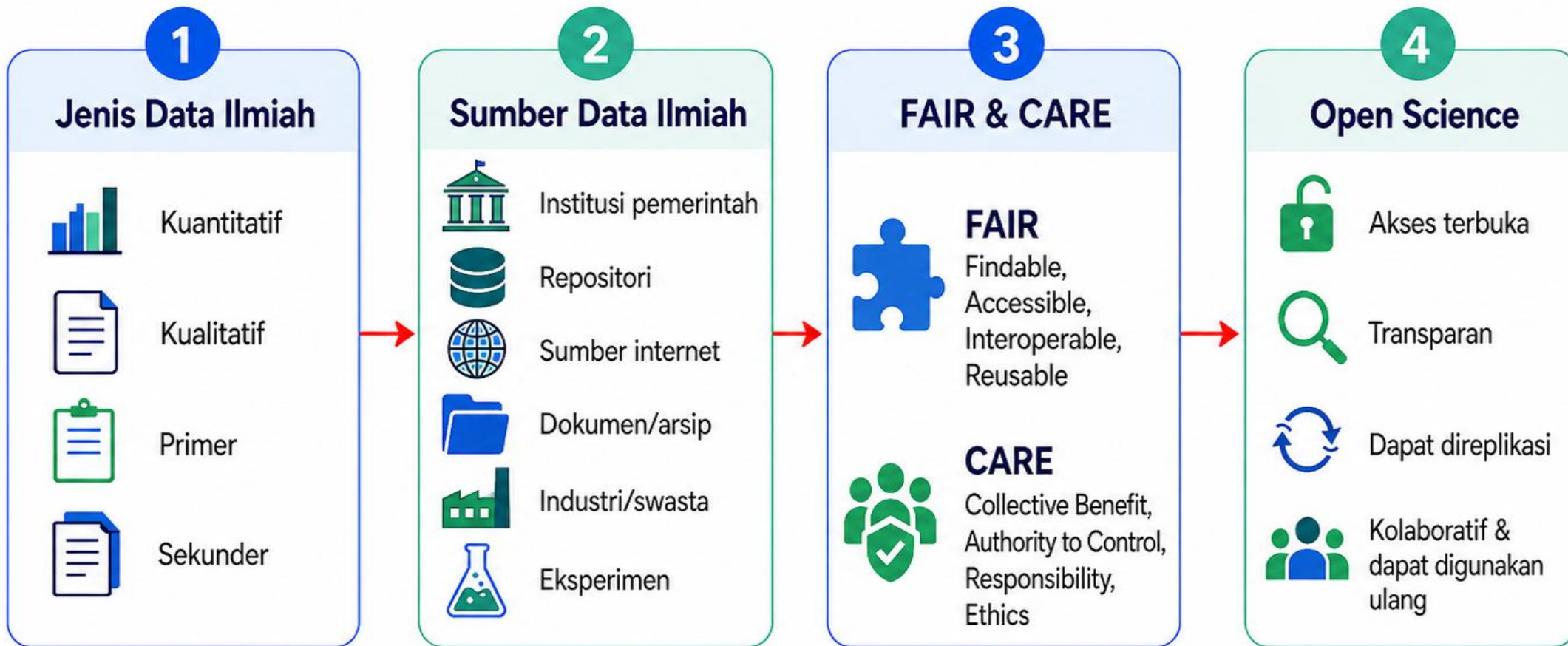
**GUNAKAN DATA
BERKUALITAS UNTUK
HASIL TERBAIK**

PIRAMIDA DIKW

Data → Informasi → Knowledge → Wisdom



JENIS & SUMBER DATA ILMIAH



Data ilmiah dari berbagai jenis dan sumber harus dikelola dengan prinsip FAIR dan CARE agar dapat dibagikan melalui **open science** secara **transparan, kolaboratif, dan bermanfaat**.



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
Pendekatan Filosofis	Positivisme	Interpretivisme

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

KAPAN DATA UMUM MENJADI DATA ILMIAH

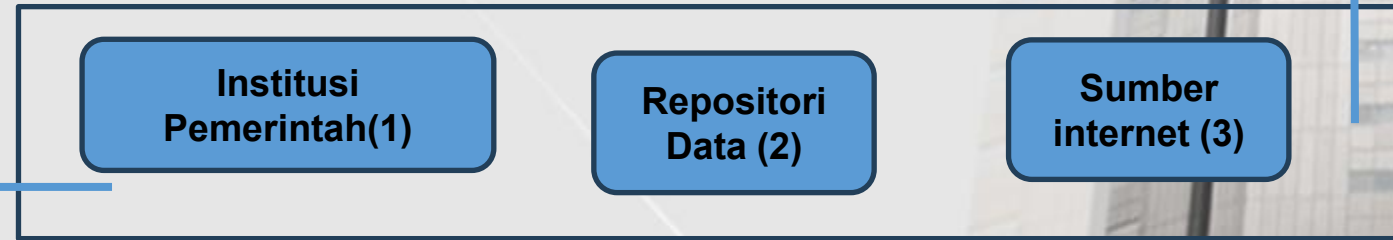
- Data umum → tidak secara khusus digunakan untuk riset



- Menjadi Data ilmiah → ada tujuan penelitian, sistematis, verifikasi, analisis
 1. cara memperoleh
 2. Memverifikasi
 3. menganalisis data

SUMBER DATA DAN INFORMASI ILMIAH

Sumber seperti Google Scholar, Altmetric.com, dan Overton digunakan untuk mengukur dampak sosial, media, dan politik dari penelitian



- ✓ Data bisnis dan survei dari pemerintah menjadi sumber penting bagi penelitian sosial
- ✓ Survei berseri dari lembaga akademik menyediakan data berkualitas tinggi untuk analisis sosial

BNPB:

https://dibi.bnpb.go.id/superset/dashboard/1/?standalone=0&expand_filters=0

BIG : <https://srgi.big.go.id/tides>

<https://siga.bkkbn.go.id/tabulasi-siga>

Repositori seperti re3data.org menyediakan akses ke berbagai database penelitian sosial humaniora

<https://www.kaggle.com/datasets?tags=11105-Education>

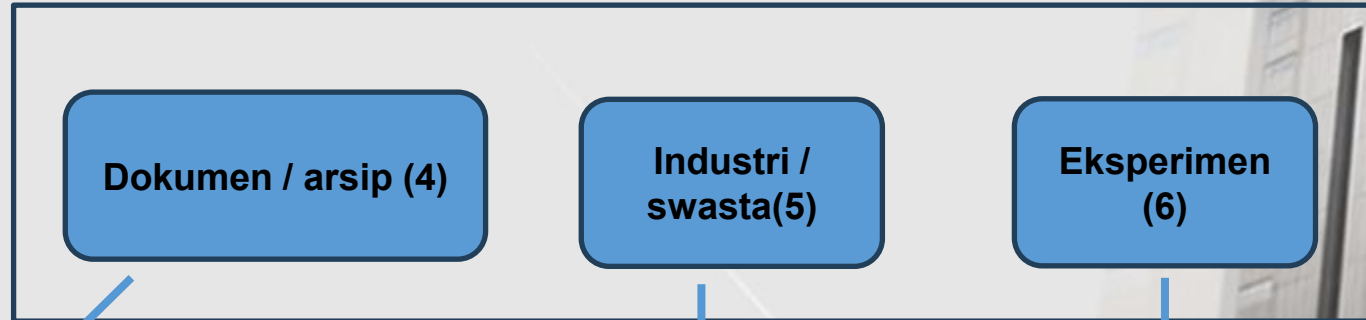
<https://www.re3data.org/search>

<https://rin.brin.go.id/>

https://datacatalog.worldbank.org/search?sort=last_updated_date%20desc

Altmetric.com
API Medsos(twitter, FB, IG, toktok dll)

SUMBER DATA DAN INFORMASI ILMIAH



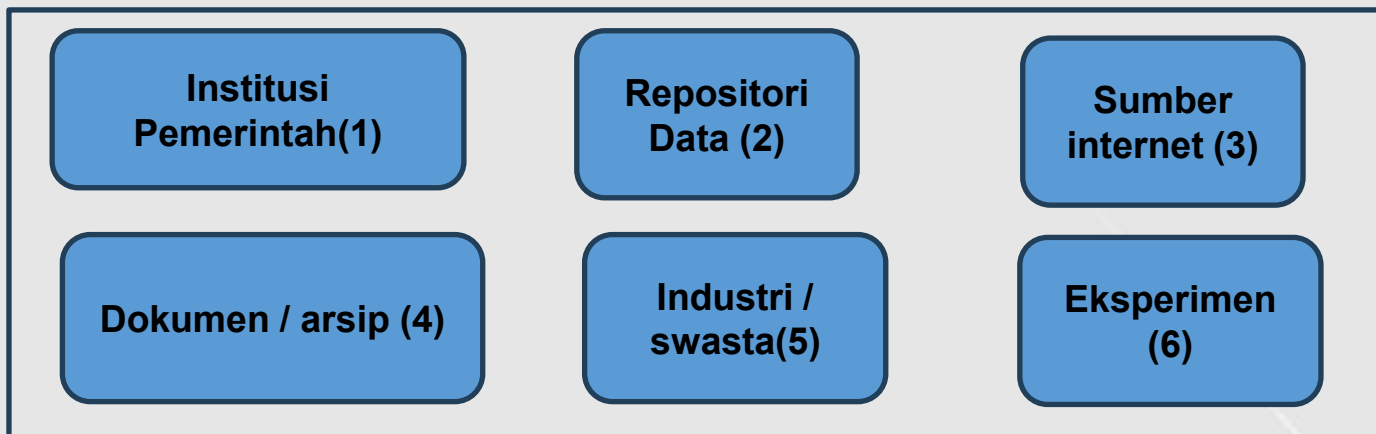
- ✓ Dokumen tradisional dan arsip sering digunakan untuk menyimpan dan menyebarkan data kualitatif
- ✓ Literatur abu-abu (grey literature) seperti ephemera, laporan internal, dan data yang tidak terdokumentasi secara formal

Data dari laporan tahunan industri
Data berbayar (Data Tempo, dll)

<https://data.tempo.co/front/datacenter/>
<https://databoks.katadata.co.id/>
<https://data.tempo.co/MajalahTeks/detail/ARM20220115192970/kami-bukan-tentara-kami-ilmuwan>

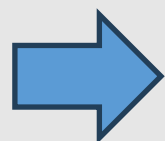
Data yang dikumpulkan melalui eksperimen laboratorium, simulasi, atau observasi langsung

SUMBER DATA DAN INFORMASI ILMIAH



Strategi
Penelusuran

FAIR ; CARE;
TRUST



sistem



Metadata



Analisis



INDIKATOR PRINSIP FAIR & CARE

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.

PRINSIP FAIR

Discipline	Application of FAIR Principles
Life Sciences	Bioimage data management, biomedical research 1 2 .
Natural Sciences	Data repositories and sharing 1 4 .
Earth Sciences	NASA Earth science data 5 .
STEM	Teaching, learning, and research workflows 6 .
Agricultural Research	FAIR and CARE principles for datasets 7 .
Social Sciences	Data storage and reproducibility 8 .
Libraries	Advocacy, training, and infrastructure 9 .
Digital Humanities	Crowdsourced collections and ethical considerations 10 .
AI and Digital Twin Systems	Addressing explainability and privacy challenges 11 .
Semantic Web	Data integration and enrichment 12 .

Discipline

Application of CARE Principles

Indigenous Studies

Indigenous Data Governance and community protocols **1**.

Library & Information Science

Research Data Management and Open Science frameworks **2**.

Data Science

Community-focused data science for social good **3**.

Health & Biomedical Sciences

Ethical health data governance and public health research **4** **5**.

Philosophy & Ethics

Ethics of care and alternative academic paradigms **6**.

METADATA

- Metadata berfungsi untuk mendeskripsikan sumber data, struktur, status, dan semantik data
- Metadata digunakan untuk mengorganisasi kumpulan data, mempermudah pencarian, dan lokasi data di masa depan
- Metadata menjadi 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

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

—

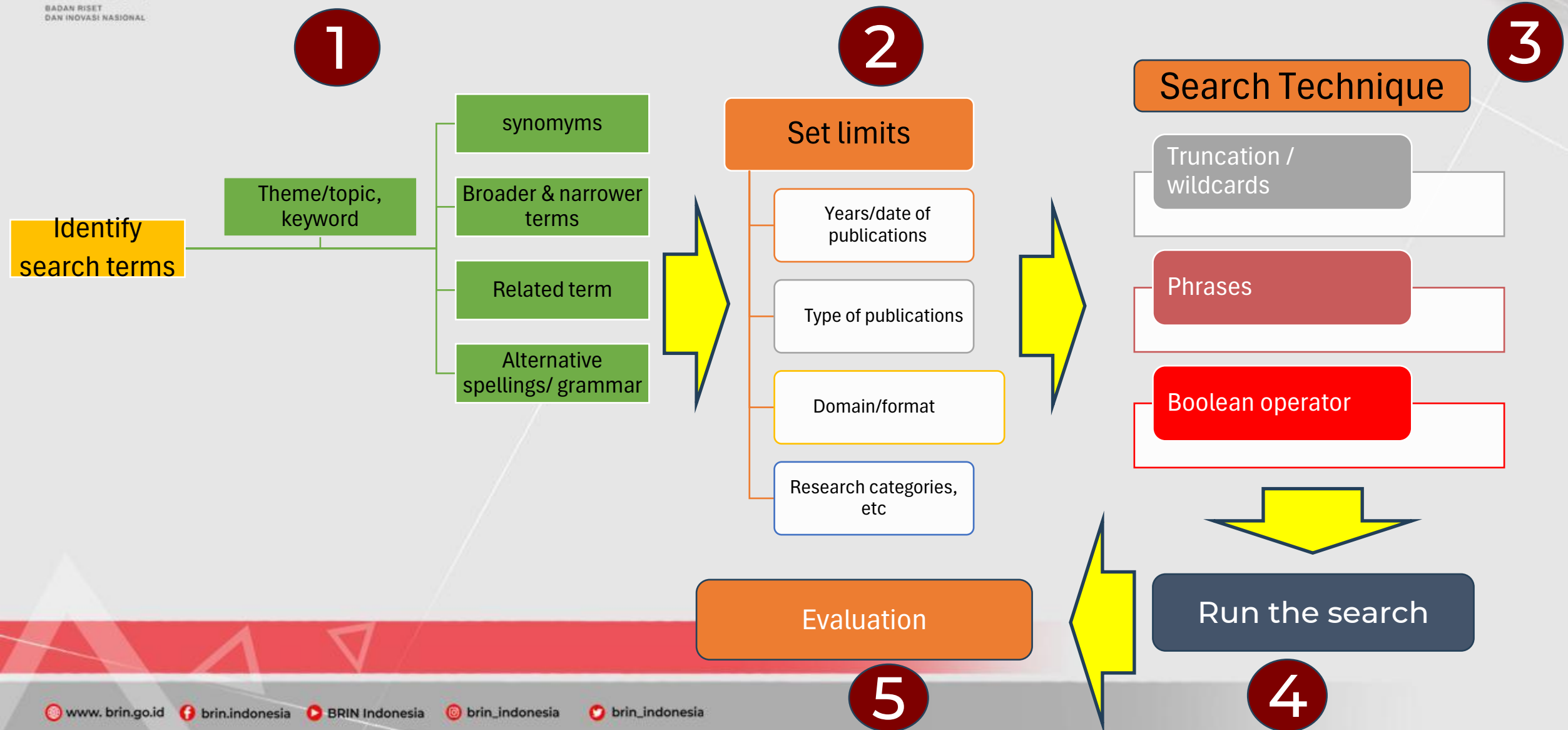
Rights

Hak cipta mengikuti ketentuan penerbit atau jurnal

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, Iftitahu (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, Iftitahu (Pusat Riset Sains Data dan Informasi BRIN)
Description 	<p>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)</p> <p>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.</p>
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, Iftitahu 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

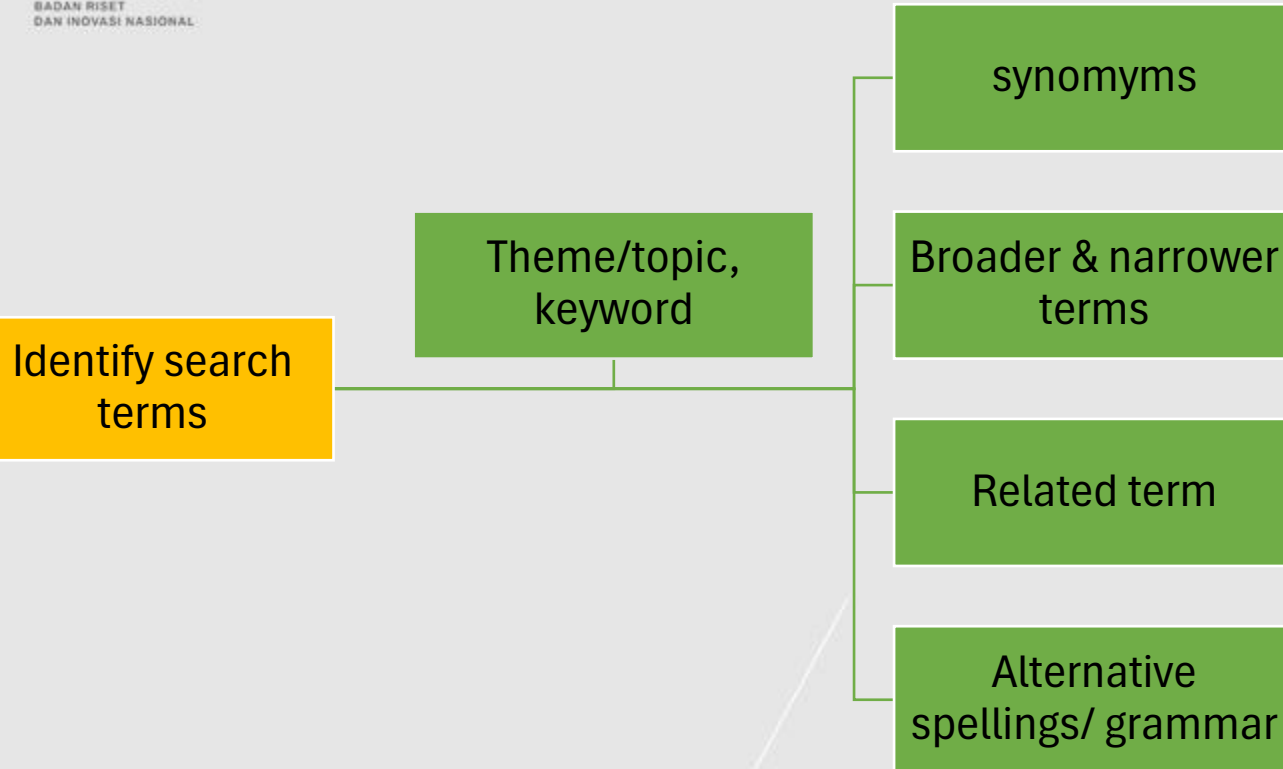


STRATEGI PENELUSURAN DATA & INFORMASI ILMIAH





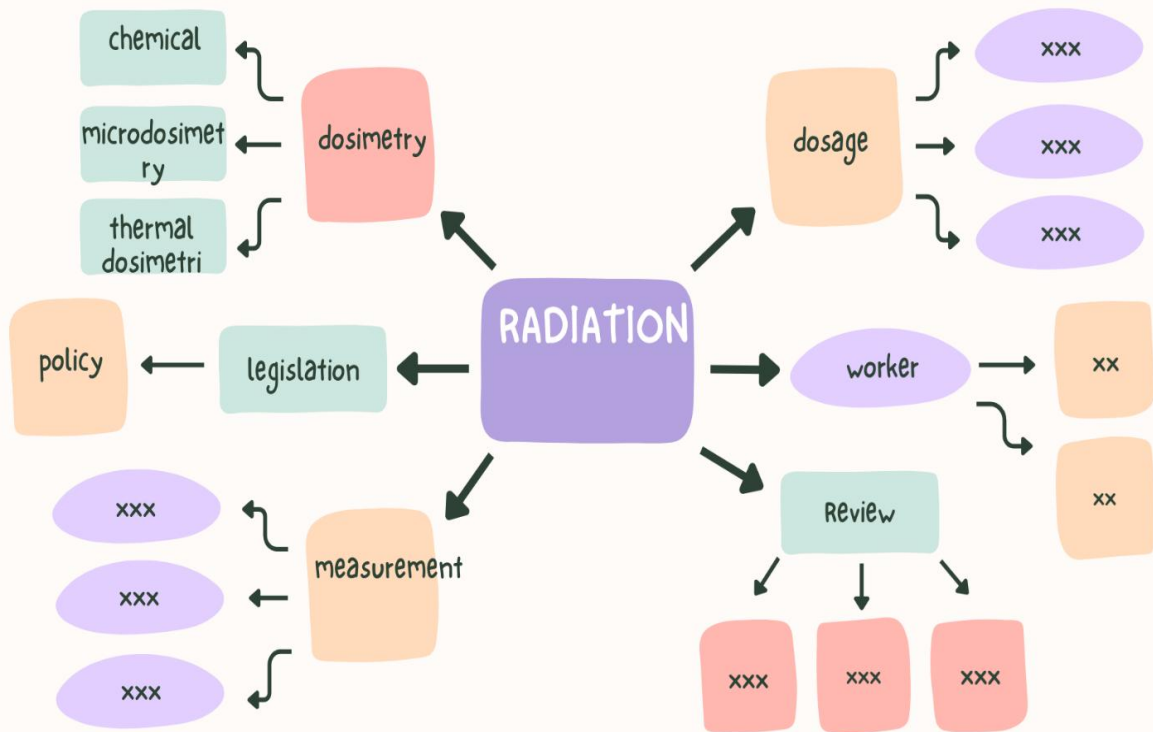
MENGIDENTIFIKASI TOPIK PENCARIAN



Tools :

1. Mind mapping
2. <https://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html>
3. <https://eric.ed.gov/?faq-thesaurus>
4. https://www.loc.gov/search?new=true&q=radiation&search_button=GO
5. <https://inis.iaea.org/search/thesaurus.aspx>
6. <https://keywordtool.io/>
7. Dictionary, etc

MENGIDENTIFIKASI TOPIK PENCARIAN



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



MENGIDENTIFIKASI TOPIK PENCARIAN

Spelling English - US	Spelling English - UK	Spelling English - US	Spelling English - UK
-e- (ex: <i>anemia, diarrhea, encyclopedia</i>)	-oe-/ae- (Ex : <i>anaemia, diarrhoea, encyclopaedia</i>)	-og (ex : <i>analog, monolog, catalog</i>)	-ogue (ex : <i>analogue, monologue, catalogue</i>)
-ense (ex: <i>defense, offense, license</i>)	-ence (ex: <i>defence, offence, licence</i>)	-ll- (ex : <i>enroll, fulfill, skillfull</i>)	-l- (ex: <i>enrol, fulfil, skilful</i>)
-el- (Ex: <i>canceled, jeweler, marvelous</i>)	-ell- (Ex: <i>cancelled, jeweller, marvellous</i>)	-o (ex: <i>color, behavior, mold</i>)	-ou (ex: <i>colour, behaviour, mould</i>)
-ize (Ex: <i>appetizer, familiarize, organize</i>)	-ise (Ex: <i>appetiser, familiarise, organise</i>)	-er (ex: <i>meter, fiber, center</i>)	-re (ex: <i>metre, fibre, centre</i>)



BRIN
BADAN RISET
DAN INOVASI NASIONAL

MENENTUKAN BATASAN

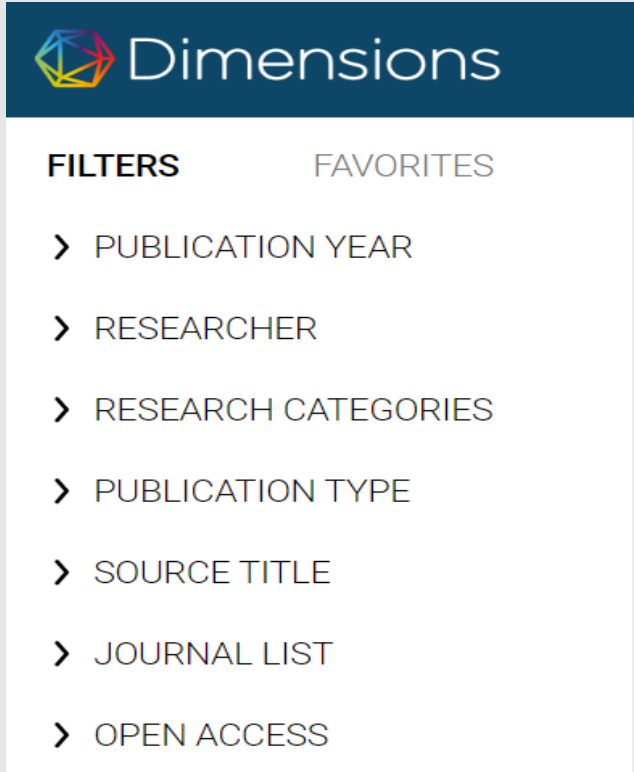
Set limits

Years/date of publications

Type of publications

Domain/format


Research categories, etc



Dimensions

FILTERS FAVORITES

- > PUBLICATION YEAR
- > RESEARCHER
- > RESEARCH CATEGORIES
- > PUBLICATION TYPE
- > SOURCE TITLE
- > JOURNAL LIST
- > OPEN ACCESS



ARTICLE TYPE

SPECIES

ARTICLE LANGUAGE

SEX

AGE

OTHER

- Address
- Autobiography
- Bibliography
- Biography
- Case Reports
- Classical Article
- Clinical Conference
- Clinical Study
- Clinical Trial Protocol

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

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



STRATEGI PENELITIAN

Search technique

Truncation /
wildcards

Phrases

Boolean operator

Truncation — a symbol added to the end of the root of a word to instruct the database to search for all forms of a word. The asterisk (*) is used in many databases for truncation.

Wildcards — a symbol used to represent any character. Wildcards can usually be used at the end of a word or within a word. The pound symbol (#) is used in many databases as a wildcard.

Search term	Word occurrence	Result
(1) behavi*	Behaviour ; behavior ; behavioral; behavior	1,172,204
(2) behavior		1,023,772
(3) wom#n	WOMN	4
(4) wom?n	Women; woman	529,828



STRATEGI PENELITIAN

Search technique

Truncation /
wildcards

Phrases

Boolean operator

Exact search or **phrase search** is used to search for information containing keywords in the form of phrases using the operator **"..." (double quote)**

Search Term	Result	note
(1) Climate change	99,330	System detect as 2 different term
(2) "Climate change"	91,572	
(3) "climate chang*"	91,739	Combine with truncation/wildcard



BRIN
BADAN RISET
DAN INOVASI NASIONAL

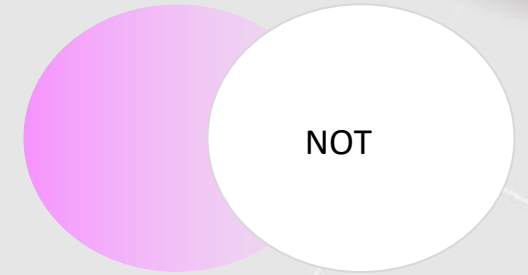
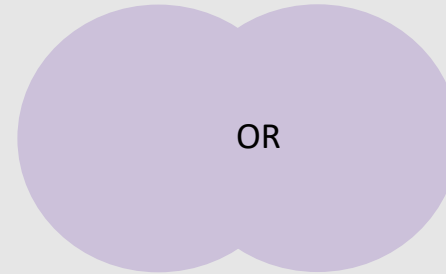
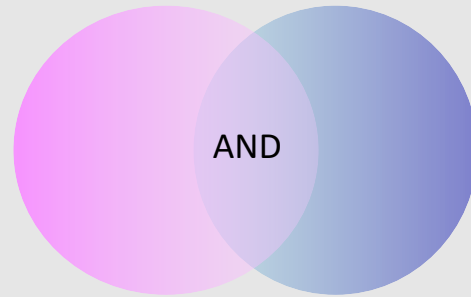
STRATEGI PENELITIAN

Search technique

Truncation /
wildcards

Phrases

Boolean operator



AND	retrieves only the middle section (intersection)
OR	retrieves all (left and right side)
NOT	excludes , only pink section is retrieved

EVALUASI PEMILIHAN SUMBER DATA DAN INFORMASI

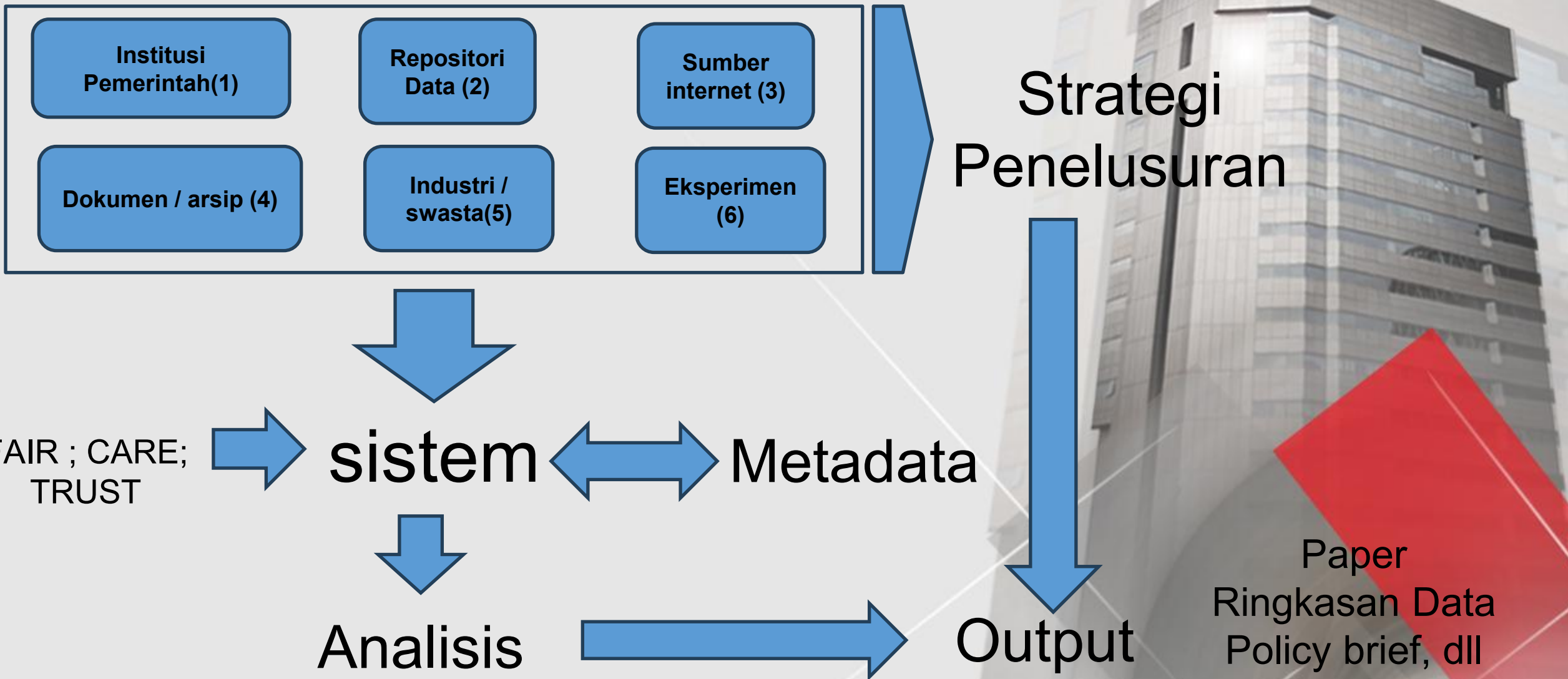
kualitas	Akurasi, relevansi, keandalan, dan konsistensi data
relevansi	Kesesuaian dengan pertanyaan penelitian
aksesibilitas	Kemudahan akses teknis dan biaya
cakupan	Wilayah geografis, waktu, atau populasi yang sesuai
kemutakhiran	Data terbaru untuk konteks yang dinamis

A framework for the quality-based selection and retrieval of open data - a use case from the maritime domain

[Stróžyna, M. opens author details in a new tab](#), [Eiden, G. opens author details in a new tab](#), [Abramowicz, W. opens author details in a new tab](#), (...), [Wećel, K. opens author details in a new tab](#)

[Electronic Markets](#) [Opens journal info in a new tab](#) 2018

SUMBER DATA & INFORMASI ILMIAH





BRIN
BADAN RISET
DAN INOVASI NASIONAL

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

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

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 Firmansyah¹; Abdurrakhman Prasetyadi²; Angga Arvianto Hadi³.

^{1,2,3} Badan Riset dan Inovasi Nasional
*Korespondensi: arvyherdianto@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://ejournal.brin.go.id/baca/issue/view/922>

Daftar Pustaka

1. Alemu, G. (2024). Metadata standards and models. In *Encyclopedia of libraries, librarianship, and information science* (1st ed., Vols. 1–4). Taylor & Francis.
2. Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE principles for Indigenous data governance. *Data Science Journal*, 19, Article 43. <https://doi.org/10.5334/dsj-2020-043>
3. Corti, L. (2007). Re-using archived qualitative data – where, how, why? *Archival Science*, 7(1), 37–54. <https://doi.org/10.1007/s10502-006-9038-y>
4. De Filippo, D., Morillo, F., & González-Albo, B. (2023). Measuring the impact and influence of scientific activity in the humanities and social sciences. *Publications*, 11(2), Article 31. <https://doi.org/10.3390/publications11020031>
5. Li, Z. (2023). A study of research data and metadata standards in the field of humanities and social sciences. In *Proceedings of the 2023 International Conference on Computer Science and Automation Technology (CSAT 2023)* (pp. 28–31). IEEE.
6. Lin, C.-S., & Wang, Y.-Y. (2018). Data type and data source preferences for six social sciences subjects in quantitative data reuses. *Proceedings of the Association for Information Science and Technology*, 55(1), 867–868. <https://doi.org/10.1002/pra2.2018.14505501151>
7. Musen, M. A., O'Connor, M. J., Schultes, E., Martínez-Romero, M., Hardi, J., & Graybeal, J. (2022). Modeling community standards for metadata as templates makes data FAIR. *Scientific Data*, 9, Article 696. <https://doi.org/10.1038/s41597-022-01815-3>
8. Putri, A. T., Chan, T.-K., Soemardi, B., & Abduh, M. (2026). Evaluating the quality of publicly available construction technology data in Indonesia. *Construction Economics and Building*, 26(1), 1–25. <https://doi.org/10.5130/h4g1we40>
9. Szomszor, M., & Adie, E. (2022). Overton: A bibliometric database of policy document citations. *Quantitative Science Studies*, 3(3), 624–650. https://doi.org/10.1162/qss_a_00204
10. Trifan, A., & Oliveira, J. L. (2020). Towards a more reproducible biomedical research environment: Endorsement and adoption of the FAIR principles. In A. Roque, A. Tomczyk, E. De Maria, F. Putze, R. Moucek, A. L. N. Fred, & H. Gamboa (Eds.), *Biomedical engineering systems and technologies: BIOSTEC 2019* (Communications in Computer and Information Science, Vol. 1211, pp. 453–470). Springer. https://doi.org/10.1007/978-3-030-46970-2_22
11. Yousefi, A., & Yousefi, S. (2007). Metadata: A new word for an old concept. *Library Philosophy and Practice*.



BRIN

BADAN RISET
DAN INOVASI NASIONAL

TERIMA KASIH