



Penerapan Data Preprocessing dan Standarisasi Database Kinerja Pendidikan pada Portal Satu Data Provinsi Sumatera Selatan

M Ezra Hilmidin Kafa, Gusmelia Testiana
Sistem Informasi, Sains Dan Teknologi, UIN Raden Fatah
ezrahilmidinkafa2018@gmail.com

Abstrak

Peraturan Presiden Nomor 39 Tahun 2019 tentang Satu Data Indonesia (SDI) mewajibkan seluruh instansi pemerintah untuk memproduksi data yang akurat, mutakhir, terpadu, dan berstandar baku agar dapat dibagipakaikan secara elektronik. Permasalahan utama yang dihadapi oleh Organisasi Perangkat Daerah, khususnya di lingkungan Dinas Pendidikan Provinsi Sumatera Selatan, adalah format pelaporan data kinerja pendidikan yang masih bersifat konvensional. Data strategis daerah seperti Angka Harapan Lama Sekolah (HLS), rekapitulasi peserta didik, dan rasio guru umumnya disajikan dalam bentuk crosstab dengan banyak penggabungan sel (*merged cells*). Hal ini menyebabkan data tidak memenuhi standar *machine-readable* untuk diintegrasikan ke dalam portal Open Data berbasis Comprehensive Knowledge Archive Network (CKAN). Penelitian ini bertujuan mengimplementasikan teknik data preprocessing dan merestrukturisasi format tabel konvensional menjadi flat table guna mendukung kelancaran publikasi metadata statistik sektoral pendidikan. Metode penelitian yang digunakan adalah deskriptif kualitatif dengan pendekatan observasi partisipatif di Bagian Perencanaan Dinas Pendidikan. Tahapan rekayasa data meliputi ekstraksi, normalisasi struktur kolom, konversi ke format Comma Separated Values (CSV), dan klasifikasi metadata pada dashboard portal Satu Data Sumsel. Hasil penelitian menunjukkan bahwa tahapan *unmerging cells* dan *filtering data anomalies* berhasil mencegah *error parsing* pada sistem importer CKAN. Pengisian kelengkapan atribut metadata secara komprehensif juga terbukti mampu mencegah ambiguitas informasi publik. Kesimpulan penelitian menegaskan bahwa kepatuhan terhadap format flat table dan kelengkapan metadata merupakan prasyarat mutlak untuk mewujudkan ekosistem tata kelola data pendidikan yang interoperabel dan transparan.

Kata kunci: CKAN, Data Preprocessing, Flat Table, Kinerja Pendidikan, Satu Data Indonesia.

Abstract

Presidential Regulation Number 39 of 2019 concerning One Data Indonesia (SDI) requires all government agencies to produce accurate, up-to-date, integrated, and standardized data for electronic interoperability. The main problem faced by Regional Apparatus Organizations, specifically within the Education Office of South Sumatra Province, is the conventional format of educational performance data reporting. Regional strategic data, such as Expected Years of Schooling (HLS), student recapitulation, and teacher ratios, are generally presented in crosstab formats with numerous merged cells. This condition causes the data to fail machine-readable standards for integration into the Comprehensive Knowledge Archive Network (CKAN) based Open Data portal. This study aims to implement data preprocessing techniques and restructure conventional table formats into flat tables to support the smooth publication of educational sectoral statistical metadata. The research method used is descriptive qualitative with a participatory observation approach in the Planning Division of the Education Office. The data engineering stages include extraction, column structure normalization, conversion to Comma Separated Values (CSV) format, and metadata classification on the Satu Data Sumsel portal dashboard. The results indicate that the stages of *unmerging cells* and *filtering data anomalies* successfully prevented parsing errors in the CKAN importer system. Comprehensive metadata attribute completion has also proven capable of preventing public information ambiguity. The study concludes that compliance with flat table formats and metadata completeness is an absolute prerequisite for realizing an interoperable and transparent educational data governance ecosystem.

Keywords: CKAN, Data Preprocessing, Flat Table, Educational Performance, One Data Indonesia

1. Pendahuluan

Perkembangan teknologi informasi dan komunikasi telah mendorong transformasi digital yang masif di berbagai sektor, termasuk penyelenggaraan tata kelola pemerintahan yang responsif dan transparan melalui optimalisasi *e-Government* (Maharani & Gamayuni, 2025). Dalam arsitektur pemerintahan digital, pemerintah dituntut untuk

mampu menghasilkan pangkalan data yang akurat, mutakhir, terpadu, dan dapat dipertanggungjawabkan guna mendukung pengambilan kebijakan berbasis bukti atau *evidence-based policy* (Saepudin et al., 2025). Untuk mewujudkan standardisasi tersebut, Pemerintah menerapkan kebijakan Satu Data Indonesia (SDI) (Presiden Republik Indonesia, 2019) yang menekankan pada manajemen data pemerintah agar bermuara pada satu ekosistem portal *Open Government Data* (OGD) secara terpusat (Zannuba et al., 2024). Penerapan konsep OGD pada portal pemerintah di berbagai daerah diharapkan mampu meningkatkan kualitas layanan administrasi dan keterbukaan informasi (Wijaya et al., 2025). Di wilayah Provinsi Sumatera Selatan, infrastruktur integrasi data lintas Perangkat Daerah diorkestrasi menggunakan *platform* repositori *open-source* berskala *enterprise* yang berbasis pada arsitektur *Comprehensive Knowledge Archive Network* (CKAN).

Dinas Pendidikan Provinsi Sumatera Selatan memiliki peran sentral sebagai produsen data utama untuk indikator kinerja pendidikan daerah. Namun, pelaksanaan kebijakan SDI dalam ekosistem pemerintahan daerah seringkali menemui hambatan teknis maupun organisasional (Azizah & Permana, 2024). Tingginya ego sektoral antar instansi sering berujung pada ketidaksinkronan data strategis dan menyulitkan tercapainya interoperabilitas (Adinegoro et al., 2025). Studi terkait implementasi kebijakan SDI di berbagai daerah juga menunjukkan bahwa belum adanya Standar Operasional Prosedur (SOP) yang memadai menyebabkan kelola pangkalan data belum berjalan dengan baik sepenuhnya (Irawan & Handayani, 2025). Permasalahan fundamental di lapangan adalah incompatibilitas struktural pada format pelaporan data internal. Laporan statistik pendidikan disusun menggunakan format *crosstab* yang dipenuhi oleh penggabungan sel (*merged cells*). Kondisi anomali ini menyebabkan kegagalan algoritma pembacaan (*error parsing*). Jika tata kelola OGD tidak dijalankan dengan manajemen basis data yang terstruktur, portal terbuka ini hanya akan menyediakan sekumpulan data yang cacat baca dan gagal dimanfaatkan (Rizky et al., 2025).

Tinjauan literatur dari penelitian terdahulu menunjukkan bahwa sebagian besar studi OGD masih bersifat evaluatif makro. Penelitian yang dilakukan oleh Sudarwono et al. (2024) dan Sa'adah & Pramono (2025) terbukti hanya berfokus pada audit penilaian tingkat kematangan menggunakan model OD-MM. Di sisi lain, penelitian oleh Putri & Testiana (2026) telah menyentuh publikasi data statistik sektoral, namun belum membedah aspek rekayasa struktur data di balik layar. *Gap analysis* dari literatur tersebut menegaskan bahwa masih sangat minim kajian teknis yang secara spesifik menawarkan solusi atas anomali format *merged cells* pada level operator daerah. Berbeda dengan penelitian sebelumnya, studi ini menawarkan *novelty* berupa penyelesaian masalah melalui teknik *data preprocessing* di level fundamental (pra-publikasi) untuk memastikan interoperabilitas basis data pendidikan. Oleh karena itu, penelitian ini bertujuan mengimplementasikan teknik *data preprocessing* untuk menstandarisasi *database* kinerja pendidikan menjadi *flat table*, serta mengoptimalkan pengisian atribut metadata pada *dashboard* CKAN Satu Data Sumatera Selatan.

2. Metode Penelitian

Penelitian ini menggunakan metode kualitatif deskriptif dengan pendekatan studi kasus dan observasi partisipatif. Pelaksanaan riset dilakukan secara luring di Bagian Perencanaan, Dinas Pendidikan Provinsi Sumatera Selatan. Dalam menelaah kolaborasi kebijakan SDI, pendekatan deskriptif sangat relevan untuk memetakan bagaimana tahapan koordinasi dari produsen data menuju publikasi informasi dilaksanakan (Anisa et al., 2023). Subjek observasi difokuskan pada praktik tata kelola *database* kinerja pendidikan, seperti data Angka Harapan Lama Sekolah (HLS) dan rekapitulasi peserta didik.

2.1. Tahapan Data Preprocessing dan Restrukturisasi

Tahapan rekayasa data diawali dengan proses Data Cleaning dan Normalisasi berbasis spreadsheet. Laporan berbasis *crosstab* dibongkar struktur hierarkinya untuk direkonstruksi menjadi format *flat table* murni. Proses pembersihan anomali pada tabel direpresentasikan melalui algoritma manipulasi berikut:

1. Tahap Ekstraksi (Data Extraction): Mengumpulkan file laporan mentah dari operator data subbagian untuk dievaluasi struktur kolomnya.
2. Tahap Eliminasi Sel Gabungan (Unmerging): Menghapus seluruh fitur format *Merged Cells* yang digunakan pada header judul dan kategori wilayah. Hal ini dilakukan agar setiap baris tabel merepresentasikan entitas tunggal (*atomic value*).
3. Tahap Pemotongan Header Ganda: Membuang hierarki header bertingkat dan baris kosong (*null records*) yang tersembunyi di dalam matriks tabel agar tidak dibaca sebagai variabel asing oleh sistem.

4. Tahap Pemfilteran Validitas: Mengaplikasikan logika filtering(String="Provinsi Sumatera Selatan") guna memastikan isolasi data hanya pada lokus administratif yang relevan.
5. Tahap Standardisasi Sintaksis: Melakukan konversi string numerik menjadi format angka murni (integer/float) dengan menghilangkan simbol mata uang atau spasi karakter.
6. Tahap Enkripsi Ekspor: Menyimpan struktur matriks ke dalam format Comma Separated Values(CSV) menggunakan protokol enkripsi UTF-8.

2.2. Manajemen Metadata dan Publikasi

Setelah instrumen CSV divalidasi keterbacaan komputasinya, tahapan bermuara pada manajemen metadata. Operator data mengakses *dashboard* portal Satu Data Sumsel berbasis CKAN menggunakan hak akses resmi institusi. Pengisian elemen metadata diklasifikasikan secara hierarkis dengan mengacu pada prinsip interoperabilitas, yang meliputi penamaan judul spesifik, penjabaran definisi operasional, penentuan organisasi produsen (Dinas Pendidikan), penetapan tahun anggaran, deklarasi lisensi terbuka publik (*Creative Commons*), serta penyisipan kata kunci pencarian (*Tags*) sektoral.

3. Hasil dan Pembahasan

Berdasarkan penerapan metodologi observasi partisipatif dan *data preprocessing*, penelitian ini berhasil mengurai kebuntuan alur digitalisasi pangkalan data di lingkungan Dinas Pendidikan Provinsi Sumatera Selatan. Terdapat empat *dataset* kinerja dan statistik pendidikan utama yang berhasil direstrukturisasi dan dipublikasikan ke dalam ekosistem portal Satu Data Sumsel, yaitu: Dataset Angka Harapan Lama Sekolah (HLS), Data Rekapitulasi Jumlah Peserta Didik, Data Rasio Ketersediaan Guru tingkat SMA/SMK/SLB, serta Dataset Kelayakan Sarana dan Prasarana Sekolah.

3.1. Restrukturisasi Flat Table dan Mitigasi Error Parsing

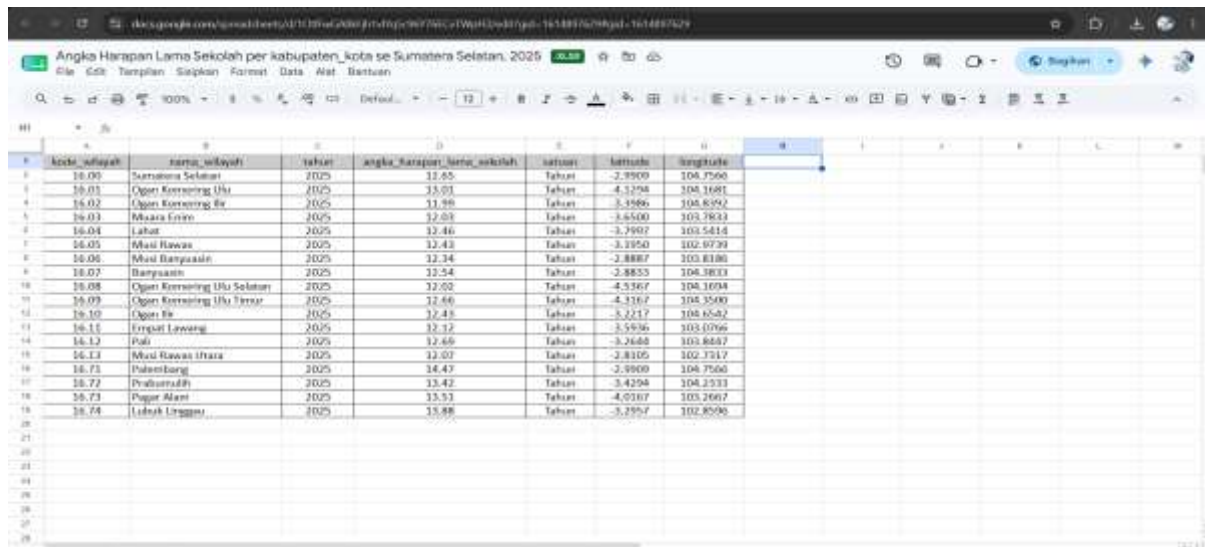
Hasil investigasi terhadap insiden kegagalan sistem mengonfirmasi bahwa penolakan API portal CKAN sepenuhnya berakar pada inkonsistensi struktur *spreadsheet* internal. Pada *file* mentahnya, staf penyusun di Dinas Pendidikan umumnya menggunakan teknik visualisasi cetak di mana nama wilayah digabungkan secara vertikal (*merge cells*). Secara komputasi kompilasi CKAN, struktur *crosstab* ini memicu generasi baris bernilai *null* atau *Not a Number* (NaN) saat diekspor ke dalam format CSV. Kegagalan *parsing* ini sejalan dengan penilaian OD-MM pada berbagai instansi bahwa lemahnya standardisasi format *machine-readable* kerap menahan tingkat kematangan *Open Data*.

Sebagai contoh, pada data rekapitulasi jumlah peserta didik tingkat SMA/SMK, kategori wilayah "Kota Palembang" seringkali di-*merge* untuk menaungi puluhan baris nama sekolah di bawahnya guna menghemat ruang visual. Secara komputasi kompilasi CKAN, struktur *crosstab* hierarkis ini memicu anomali fatal. Ketika data diekspor ke dalam format CSV, algoritma *parser* pembaca data yang bekerja secara linear dari kiri ke kanan dan atas ke bawah hanya akan membaca nilai "Kota Palembang" pada baris pertama saja. Sementara itu, puluhan baris sekolah di bawahnya akan terdeteksi sebagai sel kosong bernilai *null* atau *Not a Number* (NaN), yang secara otomatis merusak skema relasional *database*. Kegagalan *parsing* ini sejalan dengan penilaian OD-MM pada berbagai instansi yang menyimpulkan bahwa lemahnya standardisasi format *machine-readable* dari sisi hulu kerap menahan tingkat kematangan penyelenggaraan *Open Data* instansi pada level yang rendah.

Penerapan normalisasi basis data melalui *data preprocessing* bertindak sebagai solusi teknis yang absolut. Transformasi dari model relasional visual menjadi *flat table* murni menuntut prinsip bahwa satu baris wajib merepresentasikan satu *record* informasi secara independen tanpa bergantung pada sel di atas atau di sampingnya. Melalui tahapan *unmerging cells*, mahasiswa memecah sel gabungan tersebut sehingga nilai "Kota Palembang" berulang dan terisi penuh di setiap baris sekolah yang relevan. Selain itu, *header* bertingkat dipangkas menjadi satu baris judul kolom tunggal yang eksplisit (misalnya: 'Nama_Kabupaten', 'Jumlah_Siswa', 'Rasio_Guru').

Langkah terakhir dari rekayasa ini adalah memastikan *file* diekspor menggunakan format *Comma Separated Values* (CSV) dengan protokol enkripsi karakter UTF-8, guna mencegah kerusakan (*corrupt*) pada simbol atau karakter khusus. Hasilnya, algoritma ekstraktor data portal CKAN dapat membaca logika *file* secara sempurna tanpa penolakan sistem. Tidak hanya menyelesaikan masalah *error parsing*, *database* pendidikan yang telah bersih (berbentuk matriks datar) ini juga mempermudah mahasiswa dan staf di Subbagian Perencanaan dalam melakukan

olah data lanjutan, seperti menarik visualisasi infografis otomatis atau membuat bahan tayang presentasi yang interaktif menggunakan aplikasi Canva untuk kebutuhan rapat pimpinan.



kecamatan	nama_kabupaten	tahun	angka_harapan_lama_sekolah	rata_rata	persentase	tingkat
16.00	Sumatera Selatan	2025	12.65	Tahun	-2.9909	104.7560
16.01	Open Keresing Ulu	2025	13.01	Tahun	-4.3294	104.1681
16.02	Open Keresing Rie	2025	11.99	Tahun	-3.3586	104.6392
16.03	Masa Enim	2025	12.03	Tahun	-3.6500	103.7833
16.04	Lahat	2025	12.46	Tahun	-3.7902	103.5414
16.05	Musi Rawas	2025	12.43	Tahun	-3.3950	102.9739
16.06	Musi Bangsain	2025	12.34	Tahun	-2.8887	103.8180
16.07	Banyasin	2025	12.54	Tahun	-2.8853	104.3613
16.08	Open Keresing Ulu Selatan	2025	12.02	Tahun	-4.3367	104.1094
16.09	Open Keresing Ulu Besar	2025	12.46	Tahun	-4.3162	104.3580
16.10	Open Rie	2025	12.43	Tahun	-3.2217	104.6542
16.11	Open Lawang	2025	12.12	Tahun	-3.5936	103.0796
16.12	Pala	2025	12.69	Tahun	-3.2648	103.8847
16.13	Musi Rawas Utara	2025	12.07	Tahun	-2.8105	102.7317
16.14	Pulau Panglima	2025	14.47	Tahun	-2.9909	104.7560
16.15	Pulau Panglima	2025	13.42	Tahun	-3.4294	104.2333
16.16	Pagar Alam	2025	13.53	Tahun	-4.0107	103.2602
16.17	Luhuk Ulu	2025	13.88	Tahun	-3.2957	102.8096

Gambar 1. Restrukturisasi Format Tabel Angka Harapan Lama Sekolah Menjadi Flat Table

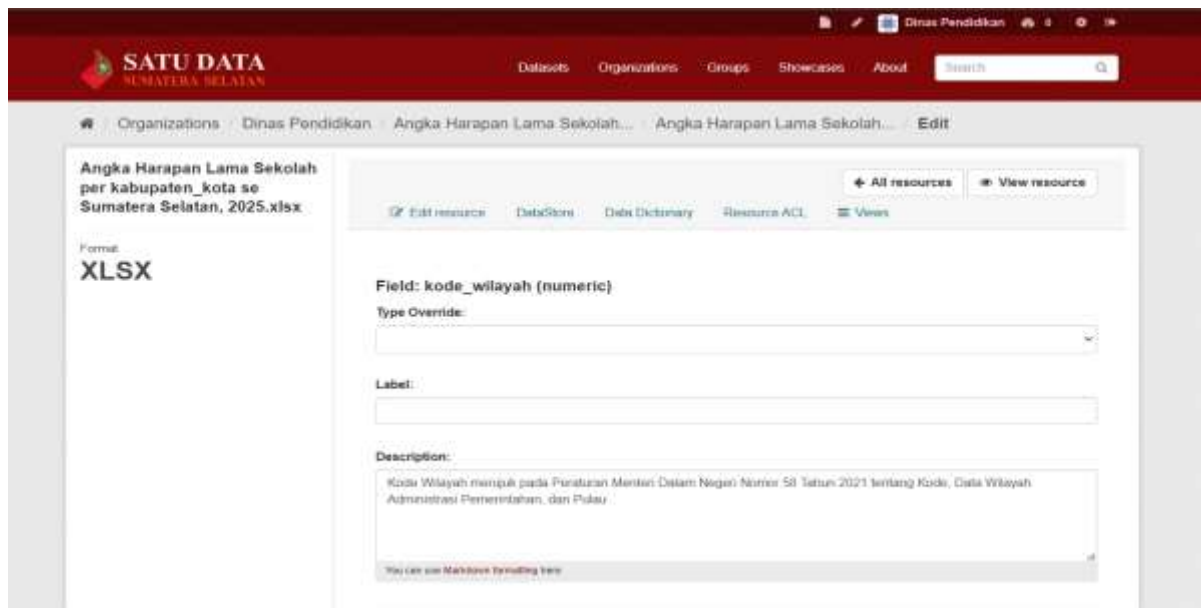
3.2. Optimalisasi Metadata dan Interoperabilitas Keamanan Data

Ketersediaan format CSV yang *machine-readable* belum memenuhi standar keterbukaan informasi yang komprehensif apabila tidak dikonfigurasi dengan deklarasi identitas metadata. Pada analisis retrospektif efektivitas SDI di Indonesia, ketiadaan metadata yang terstandar tercatat sebagai kelemahan struktural yang memicu munculnya fenomena "data yatim" (data tanpa konteks asal-usul) di dalam portal pemerintah (Maulidya & Rozikin, 2022). Observasi saat melakukan publikasi pada *dashboard* CKAN membuktikan bahwa validitas pengisian form "*Create Dataset*" secara langsung menentukan rasio keterjangkauan (*discoverability*) informasi oleh mesin pencari (*search engine*).

Proses manajemen metadata pada *dataset* kinerja pendidikan dieksekusi dengan tingkat presisi administratif yang tinggi. Sebagai contoh, alih-alih mengunggah dokumen dengan nama *file* mentah yang ambigu seperti "Rekap_Siswa_Fix.csv", mahasiswa membakukan penamaan judul (*Title*) secara profesional di portal, lengkap dengan pengisian deskripsi operasional yang menjelaskan parameter apa saja yang diukur di dalam data Angka Harapan Lama Sekolah (HLS) maupun data rasio guru. Selanjutnya, setiap *dataset* wajib mengadopsi penetapan lisensi atribusi *Creative Commons* (CC BY). Pemilihan lisensi terbuka ini sangat strategis karena menjamin hak kekayaan intelektual Pemerintah Provinsi Sumatera Selatan tetap diakui (teratribusi), namun secara bersamaan memberikan kebebasan legal bagi pihak eksternal—seperti akademisi, jurnalis data, maupun kementerian pusat—untuk menggandakan, mendistribusikan, dan menganalisis data tersebut guna kepentingan riset.

Selain itu, mahasiswa melakukan penyisipan atribut *Tags* pencarian secara spesifik, seperti kata kunci: *pendidikan*, *angka harapan lama sekolah*, *siswa*, *sarana prasarana*, dan *sumsel*. Atribut *indexing* ini memungkinkan *bot crawler* dari mesin pencari untuk memetakan dan merekomendasikan *dataset* ini secara instan ketika masyarakat melakukan pencarian terkait pendidikan di Sumatera Selatan.

Lebih jauh, pengelolaan basis data *flat table* yang bersih yang diiringi dengan jejak metadata komprehensif di ekosistem CKAN berkontribusi besar terhadap mitigasi risiko anomali siber. Di tengah tingginya ancaman *ransomware* yang menargetkan pusat data instansi pemerintah, memastikan data telah terstruktur pada *cloud* publik bertindak sebagai cadangan data (Prayugah et al., 2024). Implementasi Satu Data yang dikelola dengan komitmen kematangan infrastruktur dan keamanan yang tinggi pada akhirnya akan bermuara pada peningkatan kepercayaan masyarakat terhadap integritas penyelenggara birokrasi pemerintahan (Pramono et al., 2026).



Gambar 2. Proses Pengisian Atribut Metadata Statistik pada Aplikasi CKAN

3.3. Kualitas Data, Konsistensi Struktur, dan Kesiapan Integrasi

Hasil restrukturisasi data memperlihatkan bahwa persoalan utama pada publikasi data pendidikan tidak semata-mata berada pada tahap unggah ke portal, tetapi berawal dari cara data dibentuk sejak pertama kali dicatat. Format crosstab dengan sel gabungan memang mudah dibaca secara visual oleh operator ketika digunakan sebagai dokumen kerja atau bahan cetak, namun struktur tersebut tidak memenuhi prinsip data tabular yang konsisten untuk pemrosesan otomatis. Ketika satu nilai wilayah hanya ditulis sekali pada sel yang digabung secara vertikal, baris-baris di bawahnya kehilangan atribut wilayah secara eksplisit. Kondisi ini menyebabkan sistem membaca sejumlah record sebagai data dengan nilai kosong. Oleh sebab itu, perubahan menuju flat table bukan hanya perubahan tampilan, melainkan perubahan logika penyimpanan yang menempatkan setiap baris sebagai satu record yang lengkap dan berdiri sendiri.

Penerapan unmerging, penghapusan header bertingkat, penanganan baris kosong, dan standardisasi tipe data memberi dampak langsung terhadap kualitas basis data. Data yang sebelumnya bergantung pada struktur visual berubah menjadi data yang lebih konsisten untuk dibaca oleh manusia maupun mesin. Dalam konteks Satu Data Indonesia, kondisi ini penting karena interoperabilitas menuntut adanya struktur data yang dapat dipertukarkan tanpa kehilangan makna. Temuan penelitian ini sejalan dengan pembahasan Zannuba et al. (2024) mengenai pentingnya konsistensi tata kelola data pada implementasi Satu Data Indonesia di tingkat pemerintah daerah. Hasil tersebut juga memperkuat pandangan Saepudin et al. (2025) bahwa keterbukaan data publik akan lebih bermanfaat ketika kualitas dan aksesibilitas data telah dipastikan sejak tahap pengelolaan awal.

Dari sisi operasional, struktur flat table memudahkan proses validasi karena kesalahan dapat ditemukan pada level baris dan kolom yang jelas. Operator dapat menelusuri nilai yang kosong, mendeteksi duplikasi, memeriksa kesesuaian format angka, dan menguji konsistensi kode wilayah secara lebih sederhana. Sebaliknya, pada format crosstab, kesalahan sering tersembunyi di balik tampilan yang terlihat rapi secara visual. Dengan demikian, hasil penelitian menunjukkan bahwa data preprocessing berfungsi sebagai mekanisme kontrol kualitas sebelum data masuk ke ekosistem portal. Proses tersebut penting untuk mengurangi risiko terjadinya error parsing, salah interpretasi, dan kebutuhan perbaikan berulang setelah data telanjur dipublikasikan.

Kualitas data juga berhubungan dengan keberlanjutan pemanfaatan dataset. Ketika struktur data telah seragam, dataset dapat digunakan kembali untuk analisis statistik, visualisasi, pemetaan, maupun integrasi dengan aplikasi lain tanpa harus melakukan pembersihan ulang dari awal. Hal ini memberi nilai tambah bagi data kinerja pendidikan karena satu dataset tidak hanya berfungsi sebagai dokumen pelaporan administratif, tetapi dapat menjadi sumber informasi untuk analisis kebijakan. Praktik ini relevan dengan gagasan Open Government Data yang menempatkan data pemerintah sebagai aset publik yang dapat dipakai ulang untuk mendukung transparansi

dan pengambilan keputusan. Wijaya et al. (2025) menunjukkan bahwa keterbukaan data di pemerintah daerah dapat meningkatkan kualitas akses informasi publik apabila data tersedia dalam bentuk yang mudah ditemukan dan dimanfaatkan.

3.4. Interoperabilitas Metadata dan Kemudahan Penemuan Dataset

Selain struktur file, penelitian ini menunjukkan bahwa metadata menjadi komponen penting dalam menentukan kebermanfaatan dataset setelah dipublikasikan. File CSV yang telah bersih belum otomatis menjadi informasi publik yang mudah dipahami. Pengguna tetap membutuhkan konteks mengenai isi data, instansi produsen, periode pengukuran, cakupan wilayah, definisi variabel, serta lisensi penggunaan. Tanpa konteks tersebut, data berpotensi menjadi sekumpulan angka yang sulit ditafsirkan. Karena itu, pengisian metadata pada portal CKAN merupakan bagian yang tidak terpisahkan dari proses standarisasi data.

Penamaan dataset yang jelas dan konsisten membantu pengguna mengidentifikasi isi data sebelum membuka file. Deskripsi operasional memberikan penjelasan tentang variabel yang tersedia, sedangkan tags membantu proses pencarian berdasarkan tema. Dalam penelitian ini, atribut seperti pendidikan, angka harapan lama sekolah, siswa, sarana prasarana, dan Sumsel digunakan untuk memperkuat keterjangkauan dataset. Pendekatan tersebut penting karena mesin pencari maupun fitur pencarian internal portal mengandalkan elemen metadata untuk melakukan pengindeksan. Maulidya dan Rozikin (2022) menekankan bahwa lemahnya tata kelola metadata dapat menimbulkan kondisi data tanpa konteks yang mengurangi manfaat keterbukaan data.

Metadata juga mendukung interoperabilitas antarinstansi. Ketika penamaan variabel, organisasi produsen, periode data, dan lisensi ditulis secara jelas, dataset lebih mudah dibandingkan atau digabungkan dengan sumber lain. Dalam konteks pendidikan, hal ini membuka kemungkinan penggabungan data kinerja sekolah dengan data kependudukan, wilayah, fasilitas, maupun indikator pembangunan lainnya selama terdapat kesesuaian definisi dan kode referensi. Dengan demikian, standarisasi metadata bukan sekadar kebutuhan administratif portal, tetapi merupakan prasyarat untuk membangun rantai pemanfaatan data lintas sektor.

Temuan penelitian ini dapat dibaca sebagai bentuk penguatan tahap hulu dalam implementasi Satu Data Indonesia. Berbagai penelitian sebelumnya lebih banyak menyoroti koordinasi, kematangan Open Government Data, dan tata kelola kebijakan pada tingkat organisasi. Sudarwono et al. (2024) dan Sa'adah serta Pramono (2025) menunjukkan pentingnya tingkat kematangan penyelenggaraan Open Government Data, sedangkan penelitian ini memperlihatkan dimensi teknis yang lebih rinci, yaitu bagaimana kesalahan pada struktur spreadsheet dapat memengaruhi keberhasilan publikasi data. Dengan demikian, kualitas implementasi SDI bergantung pada keterhubungan antara aspek kebijakan, sumber daya manusia, prosedur kerja, standar metadata, dan disiplin teknis pengelolaan file.

3.5. Implikasi Preprocessing terhadap Efisiensi Kerja Pengelola Data Pendidikan

Dari perspektif proses kerja, preprocessing yang dilakukan dalam penelitian ini berpotensi mengurangi pekerjaan korektif yang berulang pada saat publikasi. Apabila kesalahan struktur baru ditemukan ketika file diunggah ke CKAN, operator harus kembali membuka dokumen sumber, mencari bagian yang bermasalah, memperbaiki format, mengeksport ulang, kemudian mengulangi proses unggah. Siklus tersebut dapat terjadi berkali-kali apabila tidak terdapat pola standarisasi yang jelas. Dengan menerapkan pemeriksaan di tahap awal, kesalahan dapat ditangani sebelum data masuk ke portal sehingga alur kerja menjadi lebih terkontrol.

Efisiensi tersebut penting bagi unit perencanaan yang mengelola banyak dataset dengan periode pembaruan berbeda. Data jumlah peserta didik, rasio guru, sarana prasarana, maupun indikator HLS dapat mengalami pembaruan secara berkala. Jika setiap dataset menggunakan struktur yang berbeda, waktu operator akan banyak digunakan untuk menyesuaikan format daripada melakukan verifikasi substansi. Sebaliknya, penggunaan flat table yang konsisten memungkinkan template kerja dibakukan dan dipakai berulang pada periode berikutnya. Dalam jangka menengah, praktik ini dapat mendukung pembentukan SOP pengolahan data yang lebih stabil sebagaimana isu yang juga ditemukan dalam implementasi SDI pada sejumlah daerah (Irawan & Handayani, 2025).

Efisiensi tidak hanya berkaitan dengan kecepatan, tetapi juga dengan penurunan risiko kesalahan manusia. Struktur data yang sederhana membuat proses pemeriksaan lebih transparan karena operator dapat melihat dengan jelas apakah setiap baris telah memiliki nilai wilayah, tahun, indikator, dan satuan yang tepat. Standarisasi sintaksis

pada data numerik juga mengurangi risiko kegagalan analisis akibat campuran simbol, spasi, atau karakter nonnumerik. Tahap ini sangat relevan ketika data akan digunakan oleh aplikasi yang tidak memiliki kemampuan menafsirkan format visual seperti manusia.

Hasil penelitian juga menunjukkan bahwa data yang telah bersih lebih mudah digunakan untuk kebutuhan internal setelah proses publikasi. Unit perencanaan dapat memanfaatkan dataset yang sama untuk membuat grafik, infografis, bahan rapat, atau analisis komparatif antarwilayah. Dengan kata lain, manfaat preprocessing tidak berhenti pada keberhasilan upload ke CKAN, tetapi juga memperbaiki kualitas bahan kerja organisasi. Hal ini memperkuat posisi data sebagai sumber daya bersama yang dapat digunakan oleh berbagai fungsi, bukan hanya sebagai dokumen pelaporan satu kali.

3.6. Keterkaitan dengan Tata Kelola Satu Data Indonesia

Implementasi teknis yang ditemukan dalam penelitian ini memiliki hubungan langsung dengan prinsip Satu Data Indonesia, terutama mengenai standar data, metadata, interoperabilitas, dan penggunaan kode referensi yang konsisten. Peraturan Presiden Nomor 39 Tahun 2019 menempatkan integrasi data sebagai bagian dari tata kelola pemerintahan berbasis data. Dalam praktiknya, integrasi tersebut hanya dapat berjalan apabila produsen data mampu menyediakan data dalam format yang konsisten dan dapat diproses. Temuan berupa error parsing akibat merged cells menunjukkan bahwa kepatuhan pada prinsip SDI harus diterjemahkan sampai ke tingkat operasional pengolahan spreadsheet.

Kondisi ini memperlihatkan bahwa keberhasilan SDI bukan semata-mata persoalan tersedianya portal. Portal hanya merupakan lapisan distribusi, sedangkan kualitas informasi tetap ditentukan oleh proses produksi di instansi sumber. Jika data dari produsen masih tidak terstruktur, portal akan menerima beban koreksi yang seharusnya diselesaikan sebelumnya. Oleh karena itu, penguatan kapasitas operator menjadi bagian penting. Operator perlu memahami alasan di balik penggunaan flat table, bukan hanya mengikuti instruksi teknis, sehingga perubahan format dapat dipertahankan secara konsisten pada siklus data berikutnya.

Azizah dan Permana (2024) menunjukkan bahwa koordinasi merupakan salah satu tantangan implementasi SDI di daerah. Temuan penelitian ini melengkapi persoalan tersebut dengan menunjukkan bahwa koordinasi harus disertai kesepakatan teknis mengenai bentuk data yang akan dipertukarkan. Kesepakatan tersebut dapat diwujudkan melalui template, kamus data, penamaan kolom, format tanggal, tipe data, kode wilayah, dan aturan pengisian nilai kosong. Dengan adanya aturan yang konsisten, proses integrasi antarunit akan menjadi lebih sederhana karena masing-masing produsen data memiliki acuan yang sama.

Selain itu, publikasi data pendidikan membutuhkan keseimbangan antara keterbukaan dan pengendalian kualitas. Tidak semua data yang dimiliki organisasi dapat dipublikasikan tanpa proses seleksi. Data perlu diperiksa dari sisi kelengkapan, keakuratan, konteks, serta potensi informasi sensitif sebelum menjadi data terbuka. Dalam penelitian ini, fokus publikasi berada pada data statistik sektoral dan indikator kinerja yang bersifat agregat. Pendekatan tersebut sesuai dengan semangat keterbukaan informasi sekaligus memungkinkan instansi mempertahankan tata kelola yang bertanggung jawab.

3.7. Kontribusi terhadap Transparansi dan Pengambilan Kebijakan Pendidikan

Data pendidikan yang terstruktur dan mudah diakses memiliki nilai strategis bagi proses perencanaan. Indikator seperti Angka Harapan Lama Sekolah, jumlah peserta didik, rasio guru, serta kelayakan sarana prasarana dapat digunakan untuk melihat kondisi pelayanan pendidikan antarwilayah. Ketika data tersedia dalam format yang dapat dianalisis, pemerintah daerah dapat melakukan perbandingan, mengidentifikasi kesenjangan, serta menyusun prioritas intervensi berdasarkan bukti. Dengan demikian, preprocessing dan standardisasi database berkontribusi secara tidak langsung terhadap kualitas proses pengambilan kebijakan.

Keterbukaan data juga meningkatkan kemungkinan keterlibatan pihak eksternal. Akademisi dapat menggunakan dataset untuk penelitian, masyarakat dapat melihat perkembangan indikator, dan media dapat menyajikan informasi secara lebih mudah. Namun, manfaat tersebut baru dapat tercapai apabila data dapat ditemukan, dibaca, dan dipahami. Inilah alasan mengapa hasil penelitian menempatkan struktur flat table dan metadata sebagai dua komponen yang saling melengkapi. Struktur memastikan data dapat diproses, sedangkan metadata memastikan data dapat dimaknai.

Rizky et al. (2025) membahas pentingnya tata kelola Open Government Data terhadap kualitas data dan informasi publik. Dalam konteks penelitian ini, tata kelola tersebut terlihat pada proses yang sangat konkret, mulai dari membersihkan spreadsheet hingga mengisi metadata pada portal. Hasil ini menunjukkan bahwa transparansi tidak hanya dibangun melalui kebijakan publikasi, tetapi melalui pekerjaan teknis yang menjamin bahwa data yang dibuka benar-benar dapat digunakan. Jika dataset gagal dibaca sistem atau tidak memiliki konteks, keterbukaan secara formal tidak otomatis menghasilkan transparansi substantif.

Dari sudut pandang kelembagaan, keberhasilan publikasi empat kelompok dataset utama dalam penelitian ini dapat menjadi model awal untuk perluasan pada dataset pendidikan lain. Prinsip yang sama dapat diterapkan pada data tenaga kependidikan, akreditasi, bantuan pendidikan, capaian pembelajaran, atau indikator layanan lainnya selama memperhatikan aturan perlindungan data dan kebutuhan agregasi. Pengembangan bertahap akan lebih efektif apabila setiap jenis data dilengkapi dengan kamus data dan prosedur validasi yang terdokumentasi.

3.8. Keberlanjutan, Risiko, dan Rekomendasi Implementasi

Keberlanjutan merupakan aspek penting karena standardisasi yang dilakukan satu kali belum menjamin kualitas data pada periode berikutnya. Ketika terjadi pergantian operator, perubahan format laporan, atau penambahan indikator, praktik pengolahan data dapat kembali ke pola lama apabila tidak didukung oleh prosedur formal. Oleh karena itu, temuan penelitian mengarah pada perlunya penyusunan SOP yang menetapkan langkah pemeriksaan sebelum publikasi. SOP dapat memuat aturan larangan merged cells pada data mentah, penggunaan satu baris header, standar tipe data, format nilai kosong, kode wilayah, penamaan kolom, dan tata cara ekspor CSV UTF-8.

Selain SOP, mekanisme validasi otomatis dapat dikembangkan sebagai langkah lanjutan. Sistem dapat memeriksa keberadaan header ganda, kolom kosong, tipe data yang tidak konsisten, karakter yang tidak sesuai, atau duplikasi record sebelum file diterima portal. Fitur pratinjau sebelum publikasi juga membantu operator melihat bagaimana data akan dibaca oleh sistem. Rekomendasi ini sejalan dengan kesimpulan penelitian yang menyarankan adanya auto-preview verifier untuk memberi peringatan dini terhadap kesalahan struktur pada file mentah.

Aspek keamanan tetap perlu diperhatikan pada proses keterbukaan data. Penyimpanan data pada infrastruktur publik tidak boleh dipahami sebagai pengganti pengelolaan keamanan internal. Data sumber, akun operator, riwayat perubahan, dan proses persetujuan tetap membutuhkan kontrol. Prayugah et al. (2024) menunjukkan bahwa ancaman siber pada sektor pemerintah menjadi isu yang perlu diperhatikan. Karena itu, pengembangan ekosistem data terbuka perlu disertai mekanisme pencadangan, pengendalian hak akses, pencatatan aktivitas, serta pemisahan antara data publik dan data internal yang mengandung informasi sensitif.

Secara keseluruhan, penelitian ini menunjukkan bahwa transformasi tata kelola data pendidikan dapat dimulai dari perbaikan yang sederhana namun mendasar pada struktur data. Perubahan dari crosstab ke flat table, penerapan preprocessing, penggunaan format CSV UTF-8, dan pengisian metadata yang konsisten membentuk fondasi teknis bagi interoperabilitas. Nilai utama penelitian bukan hanya pada keberhasilan mengunggah dataset, tetapi pada pembuktian bahwa kualitas data harus dibangun sejak hulu. Jika prosedur ini dilembagakan dan diterapkan secara berkelanjutan, portal Satu Data Sumsel dapat berfungsi lebih optimal sebagai sarana pertukaran informasi, transparansi, dan dukungan pengambilan kebijakan pendidikan.

3.9. Sintesis Hasil dan Arah Pengembangan Tata Kelola Data Pendidikan

Jika seluruh hasil dianalisis secara terpadu, proses yang dilakukan dalam penelitian membentuk satu rantai pengelolaan data dari tahap sumber sampai tahap publikasi. Tahap pertama memastikan bahwa data mentah tidak lagi bergantung pada pola visual spreadsheet, tahap kedua memastikan setiap record memiliki struktur yang lengkap, dan tahap ketiga memberikan konteks melalui metadata sebelum dataset disebarluaskan. Rantai tersebut menunjukkan bahwa keberhasilan publikasi pada portal CKAN merupakan hasil dari serangkaian kontrol kualitas, bukan hanya keberhasilan teknis menekan tombol unggah. Dengan sudut pandang ini, preprocessing menjadi bagian dari tata kelola data karena menentukan apakah informasi yang dipublikasikan memiliki struktur, konteks, dan keterlacakan yang memadai.

Temuan ini juga menunjukkan adanya hubungan erat antara kesiapan teknis dan kesiapan organisasi. Standardisasi format tidak akan bertahan apabila hanya dipahami oleh satu operator. Prosedur perlu diterjemahkan menjadi praktik bersama di unit penghasil data, diverifikasi oleh pengelola Satu Data, dan dievaluasi secara berkala ketika

terdapat perubahan indikator. Pola koordinasi semacam ini dapat mengurangi variasi format antarperiode dan antarbidang. Adinegoro et al. (2025) menyoroti tantangan kebijakan dan keterbukaan pada implementasi Satu Data Indonesia, sedangkan hasil penelitian ini memperlihatkan bahwa persoalan tersebut memiliki dimensi operasional yang sangat konkret. Keterbukaan yang konsisten memerlukan aturan kerja yang dapat diterapkan pada file sumber, bukan hanya ketentuan pada tingkat kebijakan.

Dalam pengembangan berikutnya, proses preprocessing dapat diarahkan menjadi lebih otomatis tanpa menghilangkan tahap validasi manusia. Template spreadsheet dapat dirancang dengan struktur kolom baku, validasi isian, daftar kode referensi, dan pemeriksaan nilai wajib sehingga kesalahan dapat dicegah sejak input. Pada tahap sebelum publikasi, skrip atau modul pemeriksa dapat mendeteksi sel gabungan, header ganda, kolom tanpa nama, tipe data yang tidak sesuai, duplikasi, dan nilai kosong pada atribut penting. Hasil pemeriksaan kemudian dapat ditampilkan sebagai daftar peringatan yang harus diselesaikan operator. Pendekatan ini mempertahankan fleksibilitas kerja menggunakan spreadsheet, tetapi menambahkan mekanisme quality assurance sebelum data masuk ke portal.

Arah pengembangan lainnya adalah penyusunan kamus data pendidikan yang menghubungkan nama indikator, definisi operasional, satuan, periode, sumber, penanggung jawab, dan kode referensi. Kamus data dapat menjadi penghubung antara kebutuhan kebijakan dan implementasi teknis karena setiap kolom pada dataset memiliki definisi yang seragam. Jika kamus tersebut digunakan secara konsisten, proses integrasi antarperiode dan antarsumber akan lebih mudah dilakukan. Pada akhirnya, penelitian ini menegaskan bahwa peningkatan kualitas Satu Data Sumsel tidak selalu harus dimulai dari teknologi yang kompleks. Perbaikan struktur tabel, disiplin metadata, dokumentasi prosedur, dan penguatan kapasitas operator merupakan langkah mendasar yang dapat memberikan dampak langsung terhadap interoperabilitas, transparansi, dan pemanfaatan data kinerja pendidikan.

4. Kesimpulan

Implementasi sistem portal *Open Government Data* berbasis CKAN mensyaratkan kepatuhan teknis yang mutlak terhadap format basis data tingkat lanjut yang memiliki karakteristik ramah mesin (*machine-readable*). Melalui penerapan teknik *data preprocessing* dan *data cleaning*, dokumen pelaporan kinerja sektoral pendidikan milik Dinas Pendidikan sukses dibongkar dan direstrukturisasi menjadi wujud *flat table* murni berekstensi CSV. Modifikasi struktural ini terbukti menjadi solusi penyelesaian esensial agar instrumen evaluasi makro pendidikan (seperti Angka Harapan Lama Sekolah) dapat diurai dan dirender tanpa menghasilkan notifikasi galat pada antarmuka publik aplikasi CKAN. Lebih lanjut, integrasi kelengkapan komponen atribut metadata yang mencakup deklarasi *Tags*, penetapan lisensi atribusi, dan deskripsi variabel pada *dashboard* portal Satu Data Sumsel terbukti sangat fundamental dalam mencegah anomali ambiguitas informasi serta memastikan terwujudnya interoperabilitas aset informasi digital. Untuk penyempurnaan sistem *e-government* ke depan, disarankan agar instansi membakukan Standar Operasional Prosedur (SOP) penyusunan laporan berformat *flat table* sejak awal pengumpulan data, serta mengembangkan fitur verifikasi pratinjau otomatis (*auto-preview verifier*) pada infrastruktur arsitektur SDI untuk memberikan peringatan dini atas keberadaan sel gabungan pada *file* mentah sebelum publikasi dieksekusi.

Reference

1. ADDIN Mendeley Bibliography CSL_BIBLIOGRAPHY Adinegoro, B., Ruhuputy, M. F. A., Pambudi, L., & Arrahman, T. (2025). KEBIJAKAN SATU DATA INDONESIA: SEBUAH ANTITESIS SEMANGAT KETERBUKAAN DAN INFORMASI PUBLIK. *Jurnal Ilmu Administrasi*, 16(1).
2. Anisa, Z. L. N., Sunu, R., & Zarkasyi, A. (2023). PROSES KOLABORASI KEBIJAKAN SATU DATA SALATIGA DALAM MENYEDIAKAN INFORMASI PUBLIK DI KOTA SALATIGA. <http://www.fisip.undip.ac.id>
3. Azizah, Y., & Permana, L. (2024). Upaya Koordinasi Satu Data Indonesia (SDI) di Kota Padang. *Jurnal Administrasi Pemerintahan Desa*, 6(1), 7 <https://doi.org/10.47134/villages.v6i1.173>
4. Irawan, S., & Handayani, R. (2025). Analisis Implementasi Kebijakan Satu Data Indonesia (SDI) di Pemerintah Provinsi Banten. *PANDITA: Interdisciplinary Journal of Public Affairs*, 8(1), 1-16. <https://doi.org/10.61332/ijpa.v8i1.225>
5. Maharani, A. Q., & Gamayuni, R. R. (2025). Pengaruh E-Government dan Transparansi terhadap Kualitas Laporan Keuangan Pemerintah Daerah. *Ekonomi Keuangan Syariah Dan Akuntansi Pajak*, 2(3), 169-183. <https://doi.org/10.61132/eksap.v2i3.1>
6. Maulidya, R., & Rozikin, M. (2022). ANALISIS RETROSPEKTIF KEBIJAKAN SATU DATA INDONESIA.
7. Pramono, S., Khafian, N., Sa'adah, S., Parastry, A., Rhamdani, R., Noor, E., Ginting, N., & Mustafa, F. A. (2026). Peningkatan Kematangan Open Government Data pada Komisi Pemilihan Umum Provinsi Banten untuk Pemilu yang Berkualitas. *JURNAL ABDIMAS INDONESIA*, 6(1).
8. Prayugah, M. I., Indahyanti, U., & Ariyanti, N. (2024). ANALISIS SENTIMEN PUBLIK PADA PEMERINTAH DALAM SERANGAN RANSOMWARE DENGAN PENDEKATAN SMOTE.
9. Putri, R., & Testiana, G. (2026). Publikasi Data Statistik Sektoral Melalui Portal Satu Data Sumsel pada Rumah Sakit Ernaldi Bahar

- Provinsi Sumatera Selatan. *Jurnal Penelitian Nusantara*, 2(7), 751-757. <https://doi.org/10.59435/menulis.v2i7.1372>
10. Rizky, M., Deliarnoor, N. A., & Nuryanto, Y. (2025). TATA KELOLA OPEN GOVERNMENT DATA PADA DATA DAN INFORMASI PUBLIK STUDI PADA DISKOMINFO KOTA BANDUNG TAHUN 2023. *JURNAL EDUCATION AND DEVELOPMENT*, 13(3), 547-554. <https://doi.org/10.37081/ed.v13i3.7173>
 11. Sa'adah, S., & Pramono, S. (2025). Open Government Data Maturity Assessment at the General Elections Commission of Indonesia Using the OD-MM Model. *Jurnal Ilmu Sosial Dan Humaniora*, 8(3). <https://jayapanguspress.penerbit.org/index.php/ganaya>
 12. Saepudin, E. A., Solehatunnisa, F., Lestari, N. A., Gunawan, R. A. V., & Fadilah, S. (2025). Implementasi Satu Data Indonesia dalam Mendukung Transparansi dan Akses Data Publik. *Jejak Digital: Jurnal Ilmiah Multidisiplin*, 1(3), 442-452. <https://doi.org/10.63822/pydk5v51>
 13. Sudarwono, D. A., Prastowo, R. D., Ruldeviyani, Y., & Widoyono, B. (2024). Analisis Tingkat Kematangan Open Government Data Menggunakan OD-MM di Pemerintah Provinsi Aceh. *Jurnal Informatika Ekonomi Bisnis*, 640-651. <https://doi.org/10.37034/infab.v6i3.988>
 14. Wijaya, K. A. S., Lukman, J. P., Dewi, S. A. N. S., & Riani, N. W. (2025). Penerapan Open Government Dalam Meningkatkan Kualitas Informasi Publik di Kota Denpasar (Kajian Denpasar Open Data). *Jurnal Ilmu Sosial Dan Humaniora*, 8(1). <https://jayapanguspress.penerbit.org/index.php/ganaya>
 15. Zannuba, L. A., Satyawan, D. S., & Wijaya, S. S. (2024). IMPLEMENTASI KEBIJAKAN SATU DATA INDONESIA PADA TINGKAT PEMERINTAH KABUPATEN BANYUMAS. *PUBLIC POLICY AND MANAGAMENT INQUIRY*, 8.
 16. Presiden Republik Indonesia. (2019). *Peraturan Presiden Nomor 39 Tahun 2019 tentang Satu Data Indonesia*. Sekretariat Negara.