Infrastruktur Research Data Management Berbasis Knowledge Graph dan LLM untuk Ekosistem Riset Berkelanjutan

PPMI STEI 2026 | KK Rekayasa Perangkat Lunak dan Pengetahuan

Wikan Danar Sunindyo (ketua) · Muhamad Koyimatu · Adi Mulyanto

Sekolah Teknik Elektro dan Informatika · Institut Teknologi Bandung

Abstrak

Riset ini mengembangkan infrastruktur pengelolaan data riset (RDM) berbasis DBRepo untuk data vulkanologi Indonesia, dengan knowledge graph dan model bahasa besar (LLM) sebagai rencana pengembangan. DBRepo telah dipasang di ITB dan digunakan untuk studi kasus awal bersama tim vulkanologi Prof. Asep. Prototipe dari tugas akhir Satriadhikara Panji Yudhistira menambahkan lima fungsi agregasi untuk sitasi data berbasis kueri. Pada enam kolom vulkanologi, penjumlahan reprodusibel (RSUM) menghasilkan satu checksum per kolom pada seluruh urutan input yang diuji.

Kata kunci · FAIR, DBRepo, sitasi data, vulkanologi, knowledge graph, LLM

Pendahuluan

Data vulkanologi meliputi pengamatan, pengukuran, dan rekaman geospasial. Metadata yang tidak seragam serta riwayat pengolahan yang terbatas menyulitkan penemuan dan pemanfaatan ulang data.

Untuk data yang berubah, sitasi perlu menunjuk versi data dan kueri yang tepat. Urutan pemrosesan bilangan floating-point dapat mengubah hasil agregasi dan checksum ketika kueri yang sama dijalankan ulang.

Studi ITB–TU Wien menindaklanjuti kerangka FAIR dengan adaptasi DBRepo untuk vulkanologi. Profil metadata dirancang untuk mendokumentasikan lokasi, waktu, satuan, asal-usul data, lisensi, dan ketentuan akses.

Metode Riset

Design science research dipadukan dengan studi kasus vulkanologi. Evaluasi domain dirancang untuk membandingkan alur DBRepo dengan praktik pengelolaan data awal.

Inventarisasi & baseline

Petakan koleksi, metadata, pengguna, dan batas akses.

↓

Desain & implementasi

Rancang profil metadata FAIR, siapkan skema dan masukkan data ke DBRepo.

↓

Evaluasi domain yang direncanakan

Bandingkan kualitas metadata, penemuan data, usability, kinerja, dan kesiapan penggunaan ulang.

Gambar 1. Alur adaptasi dan evaluasi FAIR

Arsitektur Sistem

DBRepo · Prototipe teruji

Versi data dan kueri temporal → agregasi MariaDB → verifikasi checksum.

Knowledge graph · Rancangan

Hubungkan dataset, publikasi, peneliti, dan metode melalui metadata.

LLM · Rancangan

Bantu ekstraksi metadata dan penelusuran dengan bahasa alami.

Gambar 2. Komponen infrastruktur RDM

Diskusi & Hasil

Prototipe menambahkan COUNT, MIN, MAX, SUM, dan AVG pada alur pembuatan subset DBRepo. RSUM diintegrasikan sebagai fungsi agregat MariaDB untuk SUM dan AVG.

Checksum unik dari 159 rekaman vulkanologi

Kolom	SUM	RSUM
Tinggi_ft	4	**1**
koor_x	3	**1**
koor_y	3	**1**
S_Dep	4	**1**
S_Dip	2	**1**
S_Strike	3	**1**

Pada 5.663 observasi hujan Pandayora, RSUM konsisten menghasilkan 46.924,3 mm pada empat urutan input. Uji menyeluruh pada prototipe juga menunjukkan checksum yang sama untuk respons HEAD, JSON, dan CSV.

Di MariaDB, median waktu RSUM mencapai 1,39–1,49 kali SUM bawaan. Penggunaan memori meningkat seiring jumlah nilai.

Kesimpulan

RSUM menjaga kestabilan checksum saat urutan input berubah pada data yang diuji. Luaran teknis ini didokumentasikan dalam tugas akhir Panji (2026) dan makalah yang diterima di ISDIA 2027.

Tahap berikutnya mencakup validasi profil metadata FAIR dan uji pengguna, dengan target artikel studi vulkanologi di Earth Science Informatics. Kajian GNN–LLM untuk Egyptian Informatics Journal merupakan target luaran PPMI.