

Происхождение и допуск химических данных

Курпишев Иван Борисович

Результат CP2B-B3.1

18 сентября 2026. Продолжение SE3-WEB1 и исходного KLT-MATDES v0.11. Досье подготовлено; научный допуск остаётся закрытым. Концепция проекта, KLT и KLT-RBD принадлежит Ивану Борисовичу Курпишеву. Программный аудит и подготовка текста выполнены с помощью Codex.

Проверяемый пилот

Запись `dft_3d_JVASP-114724` ранее выбрана для связи с CHEM-AR-004. Сверены наблюдения 11, 12 и 13: `band_gap` / DFT-OptB88vdW; `band_gap` / DFT-TBmBJ; `bulk_modulus` / DFT-OptB88vdW-Voigt-bulk. Проверены хеш снимка, база, запись, структура, материал, фаза, поле свойства, единица, метод и численное совпадение. Это происхождение данных внутри сохранённого комплекта, а не независимое подтверждение расчёта.

Новая находка

Во всех 49 записях замороженного снимка обнаружены три несогласования: повторяющиеся символы в `elements`; `nelements` не соответствует числу разных элементов; длина `elements_ratios` не равна длине `elements`. Согласно OPTIMADE 1.1.0 эти поля должны описывать один согласованный состав [1]. Область проверки ограничена составом и привязкой узлов; полный валидатор OPTIMADE не запускался.

Граница вывода

Результат относится к сохранённому KLT-экспорту. Не установлено, возникло ли расхождение у провайдера или при экспорте. Восстановление состава по `species` и `species_at_sites` допускается только как диагностическое предложение для полностью занятых однокомпонентных узлов. Простое удаление повторов не гарантирует правильного порядка долей. Исходные байты не исправлялись.

Договор нормализации

KLT-CalculationContext-B3-v1 хранит отдельно метод канала, общую метку функционала записи и параметры уровня записи. Поле `_jarvis_func` нельзя автоматически перенести на TVmBJ-канал. Значения `encut` и `kpoint_length_unit` не объявляются подтверждёнными параметрами каждого расчёта. Сходимость, фазовая устойчивость, температура, давление, процессное время и граница неопределённости остаются неизвестными. `last_modified` является датой записи.

Математическое основание

Пусть X - множество допустимых полных описаний, E - проверенные утверждения с указанной областью применимости. Зададим $S(E) = \{x \in X: p_e(x) \text{ для каждого } e \in E\}$. Требование G является подмножеством X . Семантический допуск требует одновременно $S(E) \neq \emptyset$ и $S(E) \subseteq G$; дополнительно должны выполняться независимые условия происхождения и лицензии. Противоречивое E не даёт допуск через пустое множество.

Лемма и контрпример

Если $E_1 \subseteq E_2$ и смысл всех предикатов фиксирован, то $S(E_2) \subseteq S(E_1)$: любое описание, удовлетворяющее большему набору условий, удовлетворяет меньшему. Если $S(E_2)$ пусто, результат - CONFLICT. Если два описания с разными температурами согласуются с одинаковыми сохранёнными метаданными, температура не идентифицируется этими метаданными. Подстановка нуля недопустима. Программа реализует конечные локальные проверки этого договора, а не решатель произвольных предикатов.

Связь с реперным методом

ARBD фиксирует текстовое основание и его происхождение. Числовой репер для свойства требует отдельно подтверждённого контекста. Хеш, проективное преобразование и гармонический профиль не восполняют отсутствующее наблюдение или закон эволюции. Этот слой уточняет границы применения метода; новых доказательств DHC_chem или физических гипотез не заявлено.

Первичные файлы и лицензия

В пилотной записи raw_files пуст. INCAR, KPOINTS, POSCAR/CONTCAR, OUTCAR/vasprun.xml и связь файлов с наблюдениями не получены. Страницы конкретной записи и условий, а также выбранные ресурсы Figshare вернули ошибку доступа 403 в этой сессии. Это не доказывает отсутствие данных. Каталог NIST указывает отдельный набор raw_files [2]. Набор JARVIS DFT Data 2018 года на Figshare помечен MIT, но является изменённым набором для matminer [3]; перенос лицензии на наш снимок не обоснован.

Проверка и преемственность

17 новых тестов проверяют подмену входов, неправильный канал, попытку повышения статуса, повторяемость и отсутствие численных строк в публичной проекции. Старый аудит содержит 37 блокиров; снято 0. Новые три класса несогласования учитываются отдельно, без сложения разнородных счётчиков. SE1-SE3, текстовые источники и база сохранены. Хеш-совпадение обнаруживает несогласованные изменения, но не удостоверяет автора или независимую предрегистрацию.

Следующая точка CP2B-B3.2

1. Установить происхождение полей состава: исходный ответ провайдера и версия экспортера. 2. Получить первичные файлы пилотного расчёта и применимые условия использования. 3. Подготовить отдельную производную редакцию состава с проверкой атомных узлов и фазовой идентичности. 4. Получить независимую KLT-структуру для сравнения, сохранив запрет formula-only MATCH. 5. Повторить допуск на новых версиях; лишь затем обсуждать обучение и реальные прогнозы.

Статусы

science=ABSTAIN; DHC_chem=OPEN-NOT-CONFIRMED; EXT-RUN=NOT_RUN; truth_layer_promotion=0; T6-RA5=NOT_ISSUED; physical_reachability=NOT_ESTABLISHED; parameter_identification=NOT_PERFORMED; real_corpus_forecast=NOT_RUN.

[1] OPTIMADE 1.1.0: elements, nelements, elements_ratios

[2] NIST JARVIS databases: dft_3d and raw_files

[3] JARVIS DFT Data, 2018, modified matminer dataset