

# Масштабируемая система управления исследовательскими данными с определяемыми пользователем типами

**В.А. Дударев<sup>1</sup>, Н.Н. Киселева<sup>2\*</sup>, А. Людвиг<sup>1</sup>**

<sup>1</sup>Рурский университет Бохума, Бохум, Германия

<sup>2</sup>Институт металлургии и материаловедения им. А.А. Байкова Российской академии наук, Москва, Россия

\*kis@imet.ac.ru

**Аннотация.** Разработана система управления исследовательскими данными MatInf, ориентированная на поддержку групп исследователей, работающих с большими объемами данных, полученными в ходе высокопроизводительных экспериментов в области неорганического материаловедения. Одной из ключевых особенностей системы является ее архитектура, обеспечивающая полную поддержку пользовательских типов данных, определяемых после развертывания. Это достигается за счет гибкой конфигурации системы, позднего связывания типов данных с веб-сервисами и интеграции механизмов валидации, извлечения и визуализации данных. В рамках работы проведен детальный анализ требований к RDMS, а также сформулированы ключевые задачи, возникшие при разработке. Для их решения применены методы, ранее не использовавшиеся в данном контексте, что составляет научную новизну работы. Для обеспечения гибкости и расширяемости система поддерживает интеграцию с внешними API-сервисами (для пользовательских форматов данных) и предоставляет доступ к данным по API, обеспечивая возможности для интеграции с другими системами, например инструментами машинного обучения. RDMS разработана на основе ASP.Net Core и реляционной СУБД Microsoft SQL Server, что гарантирует ее надежность, производительность и масштабируемость. Приведены примеры использования системы для накопления экспериментальных данных, документирования экспериментов и повышения воспроизводимости исследований. Открытая архитектура и свободное распространение системы делают ее достаточно универсальным инструментом для цифровизации исследований в области неорганического материаловедения, позволяя адаптировать платформу под различные задачи, включая поддержку новых типов данных и интеграцию с внешними аналитическими инструментами. Оригинальность разработки заключается в отсутствии свободно распространяемых альтернативных решений, способных осуществлять типизированное хранение материаловедческих данных (а значит, осуществлять поиск по количественному составу вещества), поддерживать расширяемую систему пользовательских типов и интегрироваться с произвольными форматами исследовательских документов без изменения ядра системы.

**Ключевые слова:** система управления исследовательскими данными, информационная инфраструктура материаловедения, система управления материаловедческой информацией, поиск материаловедческой информации, неорганическое материаловедение

**Для цитирования:** Дударев В.А., Киселева Н.Н., Людвиг А. Масштабируемая система управления исследовательскими данными с определяемыми пользователем типами // Прикладная информатика. 2025. Т. 20. № 4. С. 5–28. DOI: 10.37791/2687-0649-2025-20-4-5-28

# Scalable research data management system with user-defined types

V. Dudarev<sup>1</sup>, N. Kiselyova<sup>2\*</sup>, A. Ludwig<sup>1</sup>

<sup>1</sup>Ruhr University Bochum, Bochum, Germany

<sup>2</sup>A.A. Baikov Institute of Metallurgy and Materials Sciences of RAS, Moscow, Russia

\*kis@imet.ac.ru

**Abstract.** The MatInf research data management system has been developed to support teams of researchers working with large volumes of data from high-throughput experiments in the field of inorganic materials science. One of the key features of the system is its architecture, which provides full support for user-defined data types specified after deployment. This is achieved through flexible system configuration, late binding of data types to web services and integration of data validation, extraction and visualization mechanisms. As part of the work, a detailed analysis of the requirements for RDMS was carried out, which allowed the formation of the main functional principles of the system, including support for data on chemical compounds, flexible object typing, access delimitation and management of relationships between data. The developed structure of the relational database implements the information storage model with the possibility of creating new types of objects related to the subject area. Additional use of links between objects (graph structure) allows to manage effectively the relationships between experimental results, materials and methods of their synthesis. To ensure flexibility and extensibility, the system supports integration with external API services to integrate user-defined data formats and provides API access to data, providing opportunities for integration with other systems such as machine learning tools. RDMS is developed on the basis of ASP.Net Core and relational DBMS Microsoft SQL Server, which guarantees its reliability, performance and scalability. Examples of using the system to accumulate experimental data, document experiments and improve the reproducibility of research are presented. The open architecture and free distribution of the system make it a fairly universal tool for digitalization of research in the field of inorganic materials science, allowing the platform to be adapted to various tasks, including support for new data types and integration with external analytical tools. The development novelty lies in the absence of freely available alternative solutions capable of maintaining typed storage of materials science data (and thus search by quantitative composition of material), support extensible user-defined types system and integration with arbitrary formats of research documents without changing the system core.

**Keywords:** research data management system, materials science information infrastructure, materials information management system, materials information search, inorganic materials science

**For citation:** Dudarev V., Kiselyova N., Ludwig A. Scalable research data management system with user-defined types. *Prikladnaya informatika*=Journal of Applied Informatics, 2025, vol.20, no.4, pp.5-28 (in Russian). DOI: 10.37791/2687-0649-2025-20-4-5-28

## Введение

По мере своего развития всё больше отраслей деятельности человека не могут обходиться без поддержки со стороны современных информационных технологий. В некотором роде повы-

шение эффективности во многих сферах является прямым следствием применения передовых разработок для хранения больших объемов данных, их последующего анализа и дальнейшей поддержки принятия решений [4]. Особенно остро эта