

Practical Prompt Engineering for AI-Supported Metadata Processing in CRIS

Danica Zendulková¹ and Otmane Azeroual²

¹ Slovak Centre of Scientific and Technical Information, Lamačská 8/a, 840 05 Bratislava, Slovakia

² FernUniversität in Hagen (FeU), Universitätsstraße 47, 58097 Hagen, Germany

danica.zendulkova@cvtisr.sk, otmane.azeroual@fernuni-hagen.de

Abstract

The growing digitalization of research management has intensified the need for high-quality, consistent, and interoperable metadata in Current Research Information Systems (CRIS). However, metadata curation in many institutions remains a manual, time-consuming, and error-prone process, limiting scalability and data reliability. Recent advances in large language models (LLMs) offer new opportunities to support metadata processing through prompt engineering, enabling semantic classification and enrichment without the need for custom model development. This paper presents a practical approach for integrating LLM-based prompt engineering into CRIS workflows, focusing on the automated assignment of scientific classifications to research project data. Using a real-world case from the Slovak CRIS, we evaluate the performance of different AI tools and demonstrate how structured prompts combined with human validation can significantly improve efficiency and consistency in metadata preparation. The results show that prompt engineering provides a scalable and low-threshold solution for institutions lacking advanced AI capabilities. The study contributes to the emerging field of AI-supported research information management by offering a reproducible and practically applicable method for enhancing metadata quality in CRIS environments.

1 Introduction

The increasing digitalization of research and innovation systems has fundamentally transformed the way scientific information is produced, managed, and evaluated. Research institutions are confronted with rapidly growing volumes of heterogeneous data, including publications, projects, datasets, and other research outputs that must be systematically captured, structured, and made interoperable across institutional and national infrastructures. In this context, Current Research Information Systems (CRIS) have evolved into critical components of research management, enabling

institutions to aggregate and analyze research activities for reporting, evaluation, and strategic decision-making (Sivertsen, 2019; Fabre et al., 2021).

At the same time, policy developments such as Open Science, research assessment reform, and the increasing adoption of the FAIR data principles have significantly raised expectations regarding the quality, transparency, and interoperability of research metadata (Wilsdon et al., 2015; Wilkinson et al., 2016; Azeroual et al., 2022). Metadata are no longer limited to descriptive functions but have become essential for enabling discoverability, integration, and reuse of research information across distributed knowledge infrastructures. Consequently, metadata quality has emerged as a key determinant of the effectiveness and reliability of CRIS-based research ecosystems.

Despite this growing importance, the production and maintenance of metadata in CRIS environments remain largely manual and heterogeneous. Previous research has consistently identified metadata quality as a persistent challenge in digital information systems, highlighting issues such as inconsistency, incompleteness, and lack of standardization (Bruce & Hillmann, 2004; Park, 2009; Borgman, 2017). In the CRIS context, these challenges are further amplified by the diversity of data sources, institutional practices, and classification schemes. Empirical studies show that varying interpretation of classification standards and the absence of harmonized workflows contribute significantly to reduced metadata reliability and interoperability (Leiva-Mederos et al., 2017; Azeroual & Schöpfel, 2019; Zeng & Qin, 2020; Zendulková et al., 2025).

At the same time, the increasing complexity of research information and the demand for real-time analytics place additional pressure on institutions to improve metadata processing efficiency. While machine learning approaches have been proposed to automate classification and enrichment tasks, their adoption in CRIS environments remains limited due to the lack of training data, technical expertise, and computational resources required for model development and deployment (Azeroual, 2025a). This creates a gap between the need for scalable metadata management and the practical capabilities of many research institutions.

Recent advances in large language models (LLMs) offer a promising alternative to traditional machine learning approaches. LLMs have demonstrated strong capabilities in natural language understanding, semantic classification, and information extraction tasks, often outperforming human annotators in specific contexts (Gilardi et al., 2023; Gatto et al., 2023). Unlike conventional models, LLMs can be applied without domain-specific training by using prompt engineering, a technique that leverages structured textual instructions to guide model behavior (Zhang et al., 2025). This enables institutions to utilize AI-driven support for metadata processing without the need for complex system integration or dedicated AI development.

In this context, AI systems can be understood not merely as analytical tools, but as infrastructural components of research information systems, comparable to core data processing and management functionalities.

In this paper, prompt engineering is conceptualized as a practical method for embedding institutional knowledge, classification schemes, and metadata standards directly into AI-assisted workflows. By structuring prompts to reflect controlled vocabularies and domain-specific rules, LLMs can generate consistent and explainable metadata suggestions that support human curators in their tasks. This approach aligns with principles of human-centered artificial intelligence, where AI systems augment rather than replace human expertise (Shneiderman, 2020).

The objective of this study is to investigate the applicability and effectiveness of LLM-based prompt engineering for metadata processing in CRIS environments. The research is grounded in a real-world use case from the Slovak CRIS, focusing on the automated classification of research projects according to established scientific taxonomies. By evaluating different AI tools and analyzing their performance in a practical setting, the study aims to provide empirically grounded insights into the potential and limitations of this approach.

The paper contributes to the field of research information management in three ways. First, it provides a practical framework for integrating prompt engineering into CRIS workflows without

requiring advanced technical infrastructure. Second, it offers empirical evidence on the performance of LLMs in metadata classification tasks under real-world conditions. Third, it advances the discussion on AI-supported research infrastructures by demonstrating how human-in-the-loop approaches can enhance metadata quality, consistency, and scalability.

2 Related Work and Positioning

The problem of metadata quality and interoperability has been extensively studied in the context of digital libraries and research information systems. Early work by Bruce and Hillmann (2004) and Park (2009) conceptualized metadata quality as a multi-dimensional construct, including accuracy, consistency, completeness, and usability. These foundational studies highlight that metadata quality is not only a technical issue but also a socio-organizational challenge shaped by workflows, standards, and human practices. In large-scale research infrastructures, these challenges are further amplified due to heterogeneous data sources and distributed responsibilities (Borgman, 2017).

Within the CRIS domain, research has emphasized the importance of interoperability and standardization for enabling integrated research information ecosystems. Leiva-Mederos et al. (2017) propose semantic interoperability frameworks to address heterogeneity across CRIS implementations, while Azeroual and Schöpfel (2019) empirically demonstrate persistent data quality issues in institutional CRIS deployments. More recent work by Zendulková et al. (2025) and Fabre and Azeroual (2024) highlights the role of knowledge graphs and semantic integration as key enablers for next-generation CRIS infrastructures. However, these approaches often assume the availability of high-quality, standardized metadata, which remains a critical bottleneck in practice.

In parallel, the broader field of research data management has advanced principles for improving data usability and interoperability, most notably through the FAIR framework (Wilkinson et al., 2016). While FAIR provides important guidelines for making data findable, accessible, interoperable, and reusable, it does not prescribe concrete methods for generating or curating high-quality metadata in operational environments such as CRIS. Similarly, Open Science initiatives emphasize transparency and accessibility but often overlook the practical challenges of metadata production and standardization at scale (Wilsdon et al., 2015).

Efforts to improve reproducibility and standardization in computational research have also contributed relevant insights. Approaches focusing on workflow documentation, code sharing, and standardized reporting aim to enhance transparency and replicability (Stodden et al., 2018). However, these approaches primarily address research outputs rather than the upstream processes of metadata generation and classification in institutional systems.

More recently, machine learning techniques have been explored for automating metadata generation and classification tasks. While these approaches show promising results, they typically require large annotated datasets, domain-specific model training, and technical expertise, limiting their applicability in many institutional contexts (Azeroual, 2025a). As a result, their adoption in CRIS environments remains relatively limited.

The emergence of LLMs has introduced a new paradigm for AI-supported information processing. Studies by Gilardi et al. (2023) and Gatto et al. (2023) demonstrate that LLMs can achieve high performance in text classification and annotation tasks, often without task-specific training. Prompt engineering has further been identified as a key mechanism for controlling and adapting LLM behavior through structured instructions (Zhang et al., 2025). However, the application of prompt engineering in the context of CRIS and metadata workflows remains largely unexplored.

This paper positions itself at the intersection of research information management, metadata quality, and AI-supported workflows. Unlike existing approaches that rely on either manual curation

or fully automated machine learning pipelines, the proposed method introduces a hybrid, human-in-the-loop approach based on prompt engineering. It addresses a critical gap in the literature by providing a practical, low-threshold solution for improving metadata quality in CRIS environments without requiring dedicated AI infrastructure or training data.

3 Methodology

The study is based on a real-world dataset consisting of approximately 1000 research project records intended for integration into the Slovak Current Research Information System (SK CRIS). The data were provided in the form of an MS Excel file and exhibited typical data quality issues, including heterogeneous structure, incomplete fields, and inconsistencies.

A critical limitation of the dataset was the absence of the “field of science and technology,” which represents a mandatory attribute in the SK CRIS data model. This classification is essential for ensuring data validity and system integration. Without this information, the dataset could not be imported into the system.

The classification of scientific fields in SK CRIS is based on the Frascati Manual (OECD, 2015) and a nationally defined three-level code list of research and development (R&D) fields. The first level corresponds to the broad scientific domains defined in the Frascati Manual, while the second level introduces moderately differentiated subfields. The third level provides a more granular classification of scientific disciplines at the national level. For example, a classification path may follow the structure: *Technical Sciences* → *Industrial Biotechnology* → *Biomaterials*.

In practice, this classification is typically assigned by the data-providing institution or, in some cases, adjusted by funding agencies. When such information is missing, it must be reconstructed as part of data quality management processes. This task is time-consuming and requires domain expertise, particularly when dealing with large datasets.

Against this background, the study addresses the following research question: *Can widely available AI tools support the assignment of appropriate scientific classifications based on project titles and descriptions in CRIS environments?*

To answer this question, a structured experimental procedure was applied:

1. Creation of a reference dataset: A subset of project records was manually annotated by domain experts to establish a ground truth classification across all three levels of the taxonomy.
2. Evaluation of AI tools: Three widely accessible AI tools—ChatGPT, Gemini, and Perplexity—were selected and tested using identical task descriptions. The objective was to assess their ability to assign relevant classifications based on textual input.
3. Comparison and selection: The outputs generated by the AI tools were compared with the reference dataset to evaluate their performance in terms of semantic correctness and structural consistency.
4. Application to the full dataset: The selected tool was applied to the complete dataset. The results were manually reviewed and corrected where necessary before integration into the CRIS system.

The experimental setup relied on prompt engineering as the central methodological mechanism. Structured prompts were designed to incorporate classification rules, contextual instructions, and hierarchical constraints. This approach enables large language models to perform domain-specific classification tasks without the need for model training or system integration, aligning with emerging research on low-threshold AI adoption in knowledge infrastructures (Azeroual, 2025b; Zhang et al., 2025).

An example of the prompt structure used in this study is as follows: "You are a research classification assistant. Based on the following project title and description, assign the appropriate scientific classification according to the provided taxonomy. Use exactly one category from each level and strictly follow the official terminology of the code list. Ensure that the classification is hierarchically consistent."

The overall methodology follows a hybrid human-in-the-loop approach, integrating prompt-based AI classification with expert-driven validation to ensure both methodological consistency and metadata reliability. This approach reflects principles of human-centered AI, where automated processes are complemented by human oversight to maintain quality and accountability (Shneiderman, 2020).

Figure 1 illustrates the complete workflow of AI-supported metadata classification, including data preparation, prompt-based processing, AI-assisted classification, and subsequent human validation.

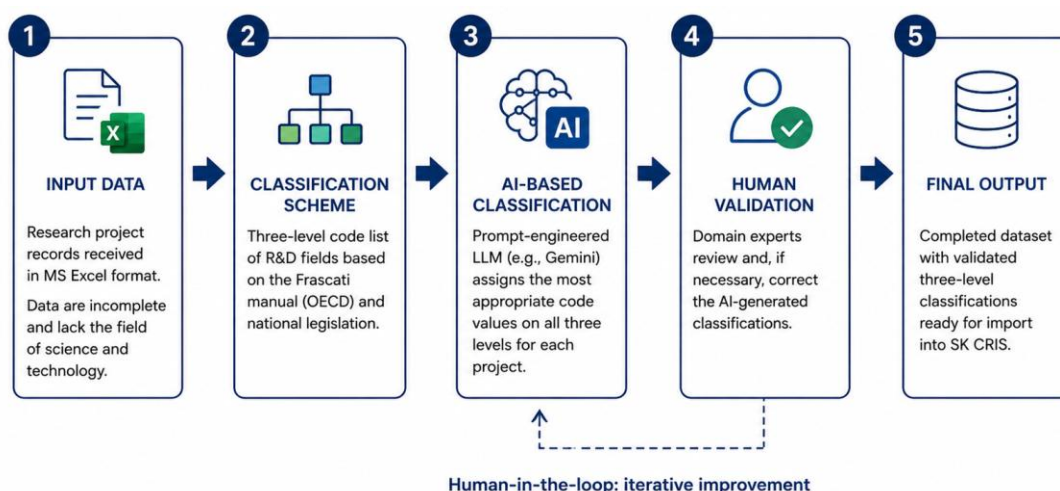


Figure 1: Workflow of AI-assisted metadata classification with human-in-the-loop validation.

The proposed approach can be conceptualized as a reusable prompt engineering framework for CRIS metadata processing. It consists of three core components: (1) structured prompt design integrating classification rules and controlled vocabularies, (2) constrained output generation aligned with hierarchical taxonomies, and (3) human-in-the-loop validation to ensure semantic and structural correctness. This framework enables institutions to systematically apply AI-supported metadata classification without requiring dedicated machine learning infrastructure.

4 Results

The experimental evaluation of AI-supported metadata classification demonstrates that LLMs can effectively support the assignment of structured scientific categories in CRIS environments. The analysis is based on a comparative assessment of three widely available AI tools—ChatGPT, Gemini, and Perplexity—applied to the same dataset and classification task. Although all tools were provided with identical instructions, their performance differed significantly in terms of semantic accuracy and structural consistency.

Perplexity was excluded at an early stage of the evaluation, as it failed to correctly apply the hierarchical classification scheme. The outputs generated by this tool were inconsistent, often including multiple unrelated categories and incorrect mappings across classification levels. This

indicates limitations in its ability to follow structured prompt instructions and adhere to domain-specific constraints. In contrast, both ChatGPT and Gemini demonstrated substantially better performance. They were generally able to infer appropriate scientific domains and assign semantically plausible categories. However, both tools initially exhibited formal inconsistencies, particularly in the precise naming of classification labels, as they attempted to approximate rather than strictly follow the official code list.

A significant improvement in classification quality was achieved by explicitly integrating the official code list into the prompt design. By providing the full taxonomy as contextual input, the selected model, Gemini, was able to align its outputs with the required hierarchical structure and terminology. This transformation of the task—from open-ended semantic interpretation to constrained classification—proved critical for achieving reliable and reproducible results. Figure 2 illustrates the structured prompt-based task assignment, in which the classification scheme is explicitly embedded and the model is instructed to assign all three levels of the taxonomy based on textual input.

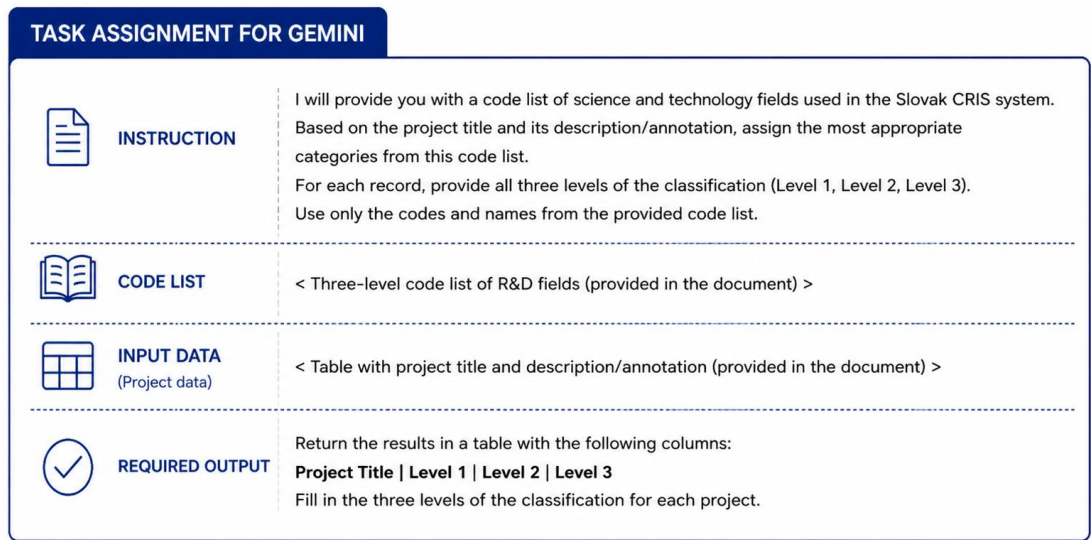


Figure 2: Task assignment for Gemini using structured prompt instructions and classification scheme.

Following tool selection, Gemini was applied to the dataset in an iterative manner. Due to technical and performance-related constraints, the data were processed in smaller batches rather than as a single input. The input consisted of project titles and corresponding descriptions, which served as the basis for semantic interpretation and classification. Although the original dataset was provided in Excel format, limitations in file handling required conversion into alternative formats, such as text or Word documents, to enable processing. Despite these constraints, the model successfully interpreted the input and generated structured outputs in tabular form.

The resulting classifications included all three hierarchical levels of the taxonomy and demonstrated a high degree of semantic alignment with the project content. Figure 3 presents an excerpt of the generated output, illustrating how project descriptions were mapped to corresponding scientific fields across multiple levels. In most cases, the model correctly identified the relevant domain and assigned consistent hierarchical classifications, particularly when the project descriptions contained sufficient contextual information.

To ensure compliance with CRIS data quality requirements, all AI-generated outputs were manually reviewed and validated by domain experts. This validation process focused on both semantic correctness and structural consistency. The majority of classifications were found to be plausible and aligned with the intended taxonomy, while necessary corrections were limited to cases

involving ambiguous or insufficient input data. These findings confirm that AI-generated metadata can significantly reduce manual workload while maintaining high quality when combined with expert validation.

No.	Project Title (Original)	Project Description / Annotation (Excerpt)	Level 1 (Field)	Level 2 (Subfield)	Level 3 (Discipline)
1	Social withdrawal in the general and clinical population: the complex relationship between social, psychological and neurobiological processes	The project examines the factors involved in social withdrawal in both the general and clinical population, including psychological, social and neurobiological mechanisms.	050000 SOCIAL SCIENCES	050100 Psychological sciences	050102 Clinical psychology
2	Pushing Boundaries in Molecular Materials by Combining Various Functionalities	Development of advanced molecular materials by combining multiple functional components to achieve novel properties and applications.	020000 ENGINEERING AND TECHNOLOGY	020500 Materials engineering	020503 Composites
3	Genetic profile variability in urinary bladder carcinoma; implications for diagnostics and therapy	Investigation of genetic variability profiles in urinary bladder carcinoma and their potential use in improved diagnostics and targeted therapy.	030000 MEDICAL AND HEALTH SCIENCES	030100 Basic medical sciences (returned from clinical medicine)	030107 Medical oncology
4	Smart energy management in buildings using IoT and machine learning	Design of an intelligent energy management system for buildings based on IoT sensors and machine learning algorithms.	020000 ENGINEERING AND TECHNOLOGY	020800 Electrical engineering, electronic engineering, information engineering	020807 Automation and control systems
5	Novel approaches to sustainable agriculture under climate change	Assessment of innovative agricultural practices and technologies to enhance sustainability and resilience under changing climate conditions.	010000 NATURAL SCIENCES	010600 Agricultural sciences	010604 Agronomy

Figure 3: AI-generated classification results with assigned three-level code values.

To complement the qualitative analysis, a subset of the dataset was quantitatively evaluated against the manually annotated reference sample. The results indicate that the model achieved a high level of accuracy on the first classification level, with slightly decreasing precision across the second and third levels due to increased granularity. The proportion of records requiring manual correction remained relatively low, confirming that the AI-assisted approach significantly reduces manual workload while maintaining acceptable classification quality.

During the evaluation, several limitations and behavioural characteristics of the system were observed. The quality of the results was sensitive to batch size, with smaller input batches yielding more accurate and stable outputs, while larger batches appeared to negatively affect model consistency. Furthermore, the performance of the model was strongly dependent on the quality of input data. Projects with vague titles or minimal descriptions could not be reliably classified, highlighting the importance of sufficiently informative metadata. Technical limitations in file processing required additional preprocessing steps, and although the use of the Slovak language did not significantly impact overall performance, it introduced minor variations in interpretation.

Overall, the results demonstrate that prompt-engineered LLMs can effectively support metadata classification tasks in CRIS environments by enabling efficient processing, consistent application of classification rules, and improved standardization of metadata. At the same time, the findings underline that such systems function most effectively as decision-support tools rather than fully autonomous solutions. The combination of structured prompts, controlled vocabularies, and human validation forms a robust hybrid approach that enhances both efficiency and reliability in metadata processing.

5 Discussion

The approach can be summarized as a four-step workflow: (1) data preparation, (2) prompt-based classification, (3) AI-assisted output generation, and (4) expert validation.

The results of this study demonstrate that prompt-engineered LLMs can significantly enhance metadata processing workflows in CRIS environments. Beyond performance improvements, the findings point to a broader transformation in how research information systems can integrate artificial intelligence as a practical and scalable component of data curation.

A central insight of this work is that the effectiveness of AI-supported metadata classification depends not only on model capabilities but also on the design of the interaction between human users and the model. In this context, prompt engineering serves as a key mechanism for translating institutional knowledge, classification rules, and controlled vocabularies into machine-interpretable instructions. This aligns with recent findings that LLMs can achieve high performance in annotation tasks when guided through structured prompts (Gilardi et al., 2023), while maintaining transparency and flexibility compared to traditional machine learning pipelines.

At the same time, the empirical observations of this study highlight several operational constraints that are particularly relevant for real-world CRIS environments. One important finding concerns the impact of batch processing on model performance. The data were processed iteratively in smaller batches, not only due to technical constraints but also as a strategy to ensure higher output quality. It was observed that smaller batches led to more accurate and consistent classifications, whereas larger input volumes resulted in a noticeable degradation of performance. This behaviour suggests limitations in context handling and may reflect internal model constraints, often perceived as a form of “model fatigue” when processing extended sequences of inputs.

Another critical factor relates to the quality and structure of the input data. The results confirm that AI-supported classification strongly depends on the availability of sufficiently informative project descriptions. Records with clear titles and detailed annotations were classified with high accuracy, while ambiguous or overly brief descriptions could not be reliably interpreted. In such cases, AI-generated outputs were either inconsistent or insufficiently precise, indicating that artificial intelligence cannot compensate for fundamentally inadequate input data. This finding reinforces existing research on metadata quality as a prerequisite for effective data processing and interoperability (Park, 2009; Zeng & Qin, 2020).

The study also highlights the continued importance of human validation within AI-supported workflows. All generated classifications were reviewed both in terms of semantic correctness and structural consistency, ensuring alignment with the hierarchical classification scheme. This validation step proved essential for identifying edge cases, such as incorrect mappings across classification levels or semantically plausible but formally invalid outputs. These observations support the view that AI should be understood as a decision-support system rather than a fully autonomous solution, consistent with principles of human-centered artificial intelligence (Shneiderman, 2020).

An additional observation concerns the role of language in AI-supported metadata processing. Although both the dataset and the classification scheme were primarily in Slovak, no significant degradation in model performance was observed. This suggests that contemporary LLMs are capable of handling multilingual classification tasks with sufficient robustness, even when operating outside English-dominated contexts. This is particularly relevant for national CRIS implementations, where local languages play a central role in research documentation.

From a methodological perspective, the study demonstrates that prompt-based approaches offer a viable alternative to traditional metadata processing methods. Existing approaches to improving metadata quality often rely on manual curation, rule-based validation, or machine learning models that require extensive training data (Leiva-Mederos et al., 2017; Azeroual & Schöpfel, 2019). In contrast, the approach presented here enables institutions to leverage AI capabilities without requiring

dedicated technical expertise or large labeled datasets. This significantly lowers the barrier to adoption and makes AI-supported metadata processing accessible to a wider range of organizations.

At a broader level, the findings suggest a shift in how AI is conceptualized within research information infrastructures. Rather than being used solely for analytical purposes, AI can function as an operational component of metadata governance and data curation processes. In this role, it contributes to improving consistency, reducing manual workload, and enhancing the overall usability of research information. At the same time, the observed limitations emphasize that effective integration requires careful design of workflows, including structured prompts, controlled vocabularies, and systematic human oversight.

Taken together, the findings of this study demonstrate that prompt-engineered AI can serve as a powerful and practical tool for improving metadata workflows in CRIS environments. However, its effectiveness depends on a combination of high-quality input data, appropriate task structuring, and continuous human validation. These findings provide important guidance for future implementations and highlight the potential of hybrid human–AI approaches in the evolution of research information systems.

The study is subject to several limitations. The evaluation is based on a single national CRIS dataset and classification scheme, which may limit the generalizability of the findings. In addition, the dataset size is moderate and does not fully capture the diversity of possible research domains. Future work should extend the evaluation to larger and more heterogeneous datasets, as well as to different CRIS implementations and multilingual environments.

6 Conclusion

The application of prompt-engineered LLMs demonstrates a practical and scalable approach to improving metadata processing in CRIS. By embedding structured prompts and controlled vocabularies into AI interactions, consistent, semantically meaningful, and reproducible classification results can be achieved without requiring complex system integration or extensive training data.

A key insight lies in the fact that the effectiveness of AI-supported metadata generation depends not only on model capabilities but on the design of human–AI interaction. Prompt engineering functions as a central mechanism for translating institutional knowledge and classification rules into machine-interpretable instructions, enabling transparency and controllability. When combined with systematic human validation, this approach ensures both efficiency gains and adherence to data quality standards required in research information management.

At the same time, several limitations must be considered. The dependency on input data quality remains a critical factor, as incomplete or ambiguous project descriptions directly affect classification accuracy. Furthermore, constraints related to batch processing and system behaviour highlight current operational boundaries of LLM-based approaches. These findings reinforce the understanding of AI as a decision-support component rather than a fully autonomous solution. The hybrid human-in-the-loop model therefore provides a robust framework for integrating AI into CRIS environments while maintaining reliability and accountability.

From a broader perspective, AI-supported metadata processing contributes to advancing interoperability, standardization, and scalability in research information infrastructures. This is particularly relevant in the context of Open Science and FAIR data principles, where high-quality metadata forms the basis for knowledge integration and reuse. Future work should focus on refining prompt strategies, exploring adaptive feedback mechanisms, and extending the approach to multilingual and cross-institutional scenarios.

These findings highlight that prompt engineering represents a low-threshold yet powerful approach for enhancing metadata workflows, offering research institutions a realistic pathway to integrate AI into existing CRIS infrastructures without significant technical overhead.

This positions AI not merely as an analytical tool, but as an operational component of research information infrastructures, fundamentally transforming how metadata is generated, validated, and maintained.

Funding

This paper was prepared with the support of project No. 401101DVH7, *Support for Slovakia's Participation in the European Research Area II (SK4ERA II)*.

References

- Azeroual, O., & Schöpfel, J. (2019). Quality issues of CRIS data: An exploratory investigation with universities from twelve countries. *Publications*, 7(1), 14.
- Azeroual, O., Schöpfel, J., Pölönen, J., & Nikiforova, A. (2022). Putting FAIR principles in the context of research information: FAIRness for CRIS and CRIS for FAIRness. In 14th International Conference on Knowledge Management and Information Systems (KMIS2022) (pp. 63-71). SCITEPRESS-Science and Technology Publications.
- Azeroual, O. (2025a). AI4Knowledge: Shaping the Future of Research Data Systems. In 2025 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA) (pp. 1-7). IEEE.
- Azeroual, O. (2025b). Streamlining Research Information Management with Large Language Models: Automation, Analysis, and User-Centric Solutions. In 10th International Conference on Big Data Analytics (ICBDA) (pp. 281-287).
- Borgman, C. L. (2017). *Big data, little data, no data: Scholarship in the networked world*. MIT press.
- Bruce, T. R., & Hillmann, D. (2004). The continuum of metadata quality: Defining, expressing, exploiting. *Proceedings of the International Conference on Dublin Core and Metadata Applications*, 1–8, ALA editions.
- Fabre, R., Egret, D., Schöpfel, J., & Azeroual, O. (2021). Evaluating the scientific impact of research infrastructures: The role of current research information systems. *Quantitative Science Studies*, 2(1), 42-64.
- Fabre, R., & Azeroual, O. (2024). Knowledge graphs—the future of integration in CRIS systems for uses of assistance to scientific reasoning. *Procedia Computer Science*, 249, 264-279.
- Gatto, J., Sharif, O., Seegmiller, P., Bohlman, P., & Preum, S. (2023, December). Text encoders lack knowledge: Leveraging generative llms for domain-specific semantic textual similarity. In *Proceedings of the Third Workshop on Natural Language Generation, Evaluation, and Metrics (GEM)* (pp. 277-288).
- Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30), e2305016120.
- Leiva-Mederos, A., Senso, J. A., Hidalgo-Delgado, Y., & Hipola, P. (2017). Working framework of semantic interoperability for CRIS with heterogeneous data sources. *Journal of Documentation*, 73(3), 481-499.

OECD. (2015). Frascati manual 2015: Guidelines for collecting and reporting data on research and experimental development, the measurement of scientific, technological and innovation activities. Paris: OECD.

Park, J. R. (2009). Metadata quality in digital repositories: A survey of the current state of the art. *Cataloging & classification quarterly*, 47(3-4), 213-228.

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*, 36(6), 495-504.

Sivertsen, G. (2019). Developing Current Research Information Systems (CRIS) as data sources for studies of research. In *Springer handbook of science and technology indicators* (pp. 667-683). Cham: Springer International Publishing.

Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences*, 115(11), 2584-2589.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., ... & Johnson, B. (2015). The metric tide. Report of the independent review of the role of metrics in research assessment and management. Higher Education Funding Council for England.

Zhang, G., Wang, X., Sun, S., Xiao, P., Sun, K., & Xiong, Y. (2025). TraceRAG: A LLM-Based Framework for Explainable Android Malware Detection and Behavior Analysis. *arXiv preprint arXiv:2509.08865*.

Zendulková, D., Azeroual, O., & Krivjanská, A. (2025). From Data Silos to FAIR Research: Interoperability and Integration in European CRIS. In *Digitalization and Management Innovation IV* (pp. 228-240). IOS Press.

Zeng, M. L., & Qin, J. (2020). *Metadata*. American Library Association.