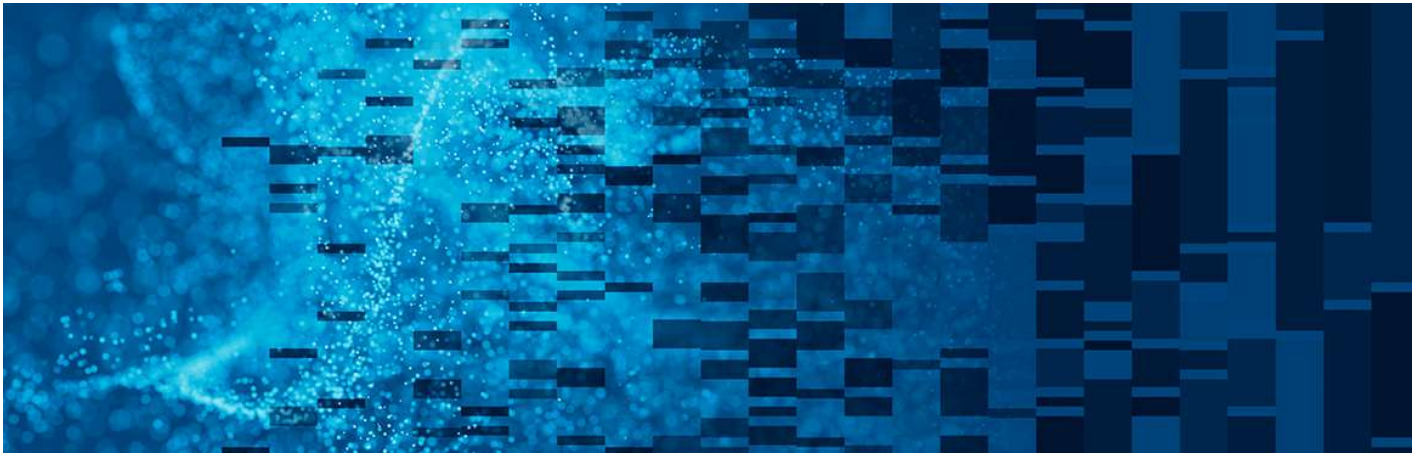


AI needs context. Linked data provides.



Axel Kaschte, Ph.D. 03 August 2026 Artificial Intelligence (AI) Linked Data Metadata Technology



Artificial intelligence is at the center of today's technology conversation. At conferences, in headlines, and across product roadmaps, AI is everywhere. At the same time, linked data, once a major focus, seems less visible.

That perception may mislead in an important way. AI is not replacing linked data. It's highlighting how essential structured, connected data really is.

These are not competing priorities. They're converging technologies that are evolving together and, in many cases, making each other better.

Understanding that relationship starts with how AI systems actually work.

AI depends on the quality of its training

Most modern AI systems, particularly those based on machine learning and neural networks, operate in two phases: training and inference.

During training, models learn from large volumes of data. These examples teach the system how patterns work, how language flows, and how relationships between concepts appear. Once trained, the model moves to the inference phase, where it applies what it has learned to new situations.

This is powerful, but it comes with challenges.

Training requires large amounts of data, and quality matters as much as quantity. If the underlying information is inconsistent, incomplete, or poorly structured, the model learns from flawed examples. When data from multiple sources is combined, those inconsistencies can multiply.

After deployment, AI systems can produce results that are difficult to verify. Hallucinations, where systems generate confident but incorrect outputs, remain a persistent challenge.

All this leads to a central question: how can we give AI better knowledge to work with?

Structured knowledge improves how AI learns

This is where linked data plays a critical role.

Linked data organizes information through structured relationships. Instead of treating data as isolated records, it connects entities and the relationships between them, forming a network of knowledge that machines can more easily interpret.

Libraries have long created this kind of structure through metadata. Formats like MARC capture rich descriptions of books, authors, subjects, and editions. As this data evolves into linked data, it becomes more than descriptive. It becomes interconnected and machine-understandable at scale.

In OCLC's experience, using structured metadata as training input can make a meaningful difference. When relationships between entities are already defined, models do not need to infer as much from raw patterns. The structure is built into the data.

This can make training more efficient and help reduce the computational effort required to reach useful results.

In simple terms, the better organized the knowledge, the easier it is for AI systems to learn from it.

Linked data also improves AI outputs

The benefits don't stop at training. Linked data also provides a way to validate AI outputs during inference.

One approach gaining attention is retrieval-augmented generation using knowledge graphs. Instead of relying solely on model output, systems can check results against a trusted source of structured knowledge.

In OCLC's own recommender work, AI-generated suggestions can be validated against WorldCat to confirm that titles exist, and the relationships are accurate before results are returned.

This adds an important layer of reliability. It reduces the likelihood of hallucinated results and provides clearer, more traceable basis for how outputs are generated.

Rather than relying only on pattern recognition, AI systems can ground their responses in verifiable knowledge.

A cycle of continuous improvement

The relationship between AI and linked data becomes even more powerful when viewed as an ongoing cycle.

AI can analyze existing metadata and identify patterns, gaps, or missing connections. Those insights can then be used to enhance the underlying data, enriching the knowledge graph.

That improved data becomes the foundation for the next round of training.

The process continues: training, inference, validation, and improvement.

At OCLC, this iterative approach is already taking shape, like with our AI deduplication process in WorldCat. Earlier this year, we completed a test run, targeting only print English books in WorldCat, and merging 5,000,000 newly detected duplicate records. Print English books represent the largest category of duplicates in WorldCat and is the format that has been most rigorously tested and improved in our machine learning deduplication activities to date. We're still analyzing the results, but one thing is clear: AI outputs can be checked against structured data, and those results can help refine and expand the data itself. Over time, both the data and the models improve together.

This is where the idea of converging technologies becomes practical. AI and linked data are not just evolving alongside one another. Each actively improves the effectiveness of the other.

Why linked data still matters

The current focus on AI can make it seem like every other technologies must adapt or fade away. The relationship between AI and linked data suggests something different.

AI is powerful, but it depends on structured, connected knowledge to reach its full potential. Linked data provides that foundation, organizing information in ways that machines can understand, validate, and build upon.

For organizations that maintain rich metadata, this represents a significant advantage. That data does more than describe resources. It enables more accurate, efficient, and trustworthy AI.

Looking ahead

AI may be the headline technology today, but it depends on reliable, well-connected data to be effective. Linked data provides that context.

At OCLC, this is not a theoretical relationship. It is an active area of investment, where linked data and AI are evolving together to improve over time.

Rather than replacing one another, these technologies are converging. And together, they are shaping how knowledge can be organized, understood, and shared more effectively.

This blog is based on the contents of the presentation “Building Intelligence: Practical library applications of AI and Linked Data at OCLC,” which was first presented at the IFLA WLIC Conference in Astana, Kazakhstan in 2025. You can watch the full recording of the presentation [here](#).