Connecting the Dots with
Linked Data

The future of library cataloging and discovery

EBSCO

MARC has a problem *(or three)...*

MARC was developed in the 1960s as a standardized cataloging format for print materials, and its age is beginning to show. While MARC may have thrived in the era of typewriters and card catalogs, the rise of the internet and digital technologies has exposed some of its fundamental limitations.

✕ MARC is **not flexible**: It is a fixed-field format, which means that each field must be used for a specific purpose. This makes it difficult to accommodate new types of information or changes in the way information is organized.

✕ MARC is **not adaptable**: While online catalogs can accommodate MARC, it is not well-suited for use in these digital environments and does not easily integrate with other systems.

✕ MARC is **not shareable**: Being primarily a local format, it's hard to share data among different systems, organizations, and services created by other industries.

Moving to linked data

Libraries are evolving to keep up with how people find and use information online, and MARC just isn't built for the digital world. Linked data offers a better way to organize and **share library metadata in the modern landscape**.

Simply put, linked data helps **identify and solidify living connections between items**. Instead of keeping data locked in local catalogs, linked data helps libraries share their knowledge more widely and link it to other sources. It works well with websites, search engines and other systems, making it easier to **establish and find connections** between things like authors, books, and subjects, and expose those connections on the web.

EBSCO

Let's look at an example of linked data:

The below web illustrates just a snapshot of the connections enabled by linked data. Follow the connections along the journey of serendipitous discovery between works, authors, subjects and more.
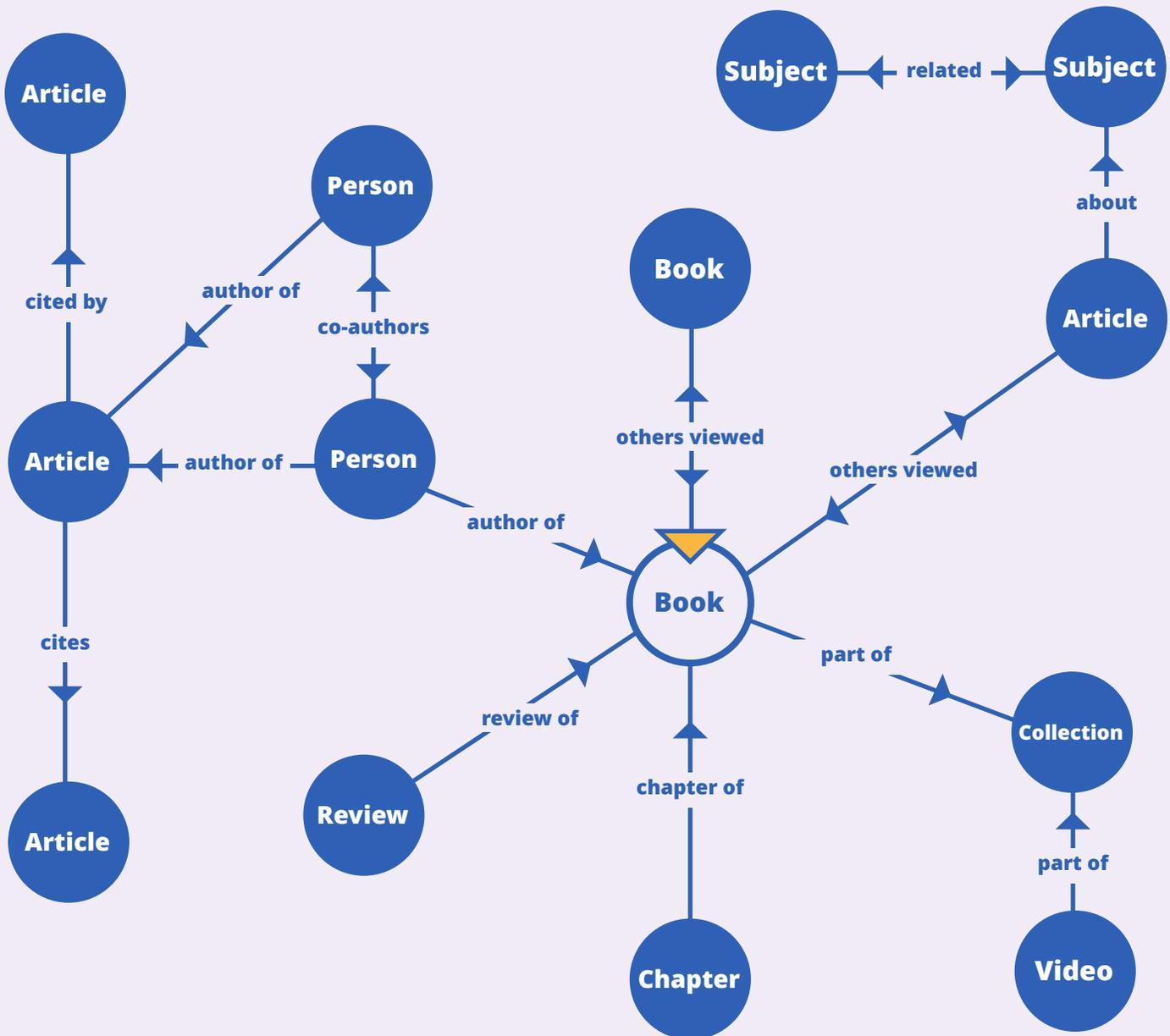
What can linked data do?

A novel set in 19th-century Paris might be linked to historical maps, photographs, and contemporaneous literature. These connections allow readers to explore the setting more vividly, perhaps discovering how the city's architecture or political climate influenced a fictional narrative. The reader may also discover real authors who lived in Paris during this time — perhaps, Victor Hugo. How did his lived experience translate into his writing, and how does it compare to the novel where we started? This kind of contextual expansion transforms a simple reading experience into a **multidimensional exploration** between fiction and nonfiction.

What can linked data do?

Or, we can imagine a researcher studying climate change policy. This researcher might discover that two scientists working on renewable energy technologies have co-authored papers with a third expert in battery storage, suggesting a potential interdisciplinary nexus worth exploring. Through linked data, they can trace how a foundational paper has been cited across disciplines — such as in economics, sociology or urban planning. This **citation trail** can reveal how the original ideas have evolved, been challenged or expanded, **helping the researcher identify emerging trends** or gaps in the literature.

EBSCO

Why does linked data matter?

✔ **Improved Discovery and Access**

Linked data enables richer search experiences by **connecting library metadata with internal and external datasets**. Linked data exposes and highlights connections between authors, works, subjects and more, allowing users to discover related resources across institutions, formats, and languages.

Furthermore, since linked data is built for the web, connections are not only machine-readable but also web-accessible, meaning your library's collections can be surfaced in broader search environments — not just within your catalog, but across the open web. This increased visibility helps users **find and engage with your resources** more easily.

✔ **Enhanced Metadata Quality**

Traditional metadata formats like MARC are rigid and often localized. Linked data, by contrast, is **flexible, dynamic and globally consistent**. It can be reused and repurposed across platforms and contexts, and authority records (e.g., for authors or subjects) are updated in real time. Shared identifiers (like VIAF or ORCID) ensure that metadata is harmonized across institutions.

This means researchers can **trust that the metadata** they encounter is accurate, current, and interoperable, reducing duplication and improving the quality of bibliographic records.

EBSCO

Why does linked data matter?

✔ ## Interoperability Across Systems

Linked data enables seamless integration between libraries, archives, museums, and other cultural heritage institutions. It connects different vocabularies and ontologies from collections across formats — from manuscripts and datasets to audio recordings and exhibitions.

This interoperability allows users to **explore diverse resources from the comfort of their trusted library platform**. For researchers, this means less time navigating systems and more time engaging with content.

✔ ## Support for Open Knowledge Initiatives

Linked data aligns with the values of open science and open scholarship by encouraging transparency in how knowledge is structured and shared, promoting accessibility to metadata and resources across institutions, and fostering collaboration between libraries through data sharing.

Libraries participating in linked data ecosystems contribute to a **collective knowledge ecosystem**, benefiting from and enriching the metadata created by others. This supports network-wide innovation, such as better recommendation systems, improved citation tracking, and more inclusive discovery tools.

Let's get technical: How does linked data work?

Vocabularies and Structure

Linked Data uses standards from the Semantic Web to make data from different sources understandable, reusable, and interoperable. These vocabularies define consistent ways to describe resources and their relationships:

- ✅ **BIBFRAME** (Bibliographic Framework): Developed by the Library of Congress to replace MARC, enabling bibliographic data to be linked and shared across systems.

- ✅ **Builde:** An openly licensed vocabulary designed by EBSCO to connect and extend linked data across communities, including cataloging, scholarly research, and archives.

- ✅ **Dublin Core:** A simple, widely adopted set of metadata terms for describing web resources like documents, images, and datasets.

- ✅ **Schema.org:** A collaborative vocabulary used across the web (including by Google and other search engines) to describe people, organizations, works, and creative content in a structured, machine-readable way.

- ✅ **SKOS** (Simple Knowledge Organization System): Used to represent controlled vocabularies such as thesauri, taxonomies, and classification schemes.

All of these vocabularies are built on RDF (Resource Description Framework), which expresses data as triples (subject, predicate, and object) enabling connections between concepts across different datasets and platforms.

EBSCO

Let's get technical: How does linked data work?

Data Governance

Linked Data helps improve data reliability and quality through data governance enforcements like SHACL (Shapes Constraint Language), which is a W3C standard used to validate RDF data against a set of constraints or "shapes." It plays a critical role in data governance by:

- Ensuring that data conforms to expected structures (e.g., a person must have a name and birthdate)
- Catching errors early in data pipelines
- Enforcing rules and semantic consistency

By applying SHACL, organizations can automate quality checks, reduce data silos, and maintain high standards of data integrity across distributed systems.

EBSCO

Let's get technical: How does linked data work?

Unique Identifiers

At the heart of Linked Data is the use of URIs (Uniform Resource Identifiers) as global UIDs (Unique Identifiers) for entities. These UIDs enable:

- ✓ Anyone to access the data behind a URI using a browser or API

- ✓ Hyperlinking on the open web to form connections between data

- ✓ Construction of knowledge graphs where each node represents a uniquely identified concept or resource

For example, a URI like **http://id.loc.gov/authorities/names/n79107741** points to structured data about Beethoven, including alternate spellings of his name, his birth and death dates, occupation, exact matches from other schemes, sources, etc. This associated data can be reused and linked to other datasets.

Speaking of knowledge graphs...

The **EBSCO Scholarly Graph** (ESG) is a structured, interconnected dataset that describes scholarly entities such as people, publications, institutions, and topics using linked data. It powers features in EBSCO products that help users **explore relationships** between authors, **discover research impact**, and **navigate academic domains** more intuitively. Built by aggregating metadata from trusted sources like OpenAlex and Crossref, the Scholarly Graph enriches this data with contextual relationships and standardized identifiers. EBSCO also calculates additional facts to expand the graph. This enables more powerful discovery experiences, such as dynamic person pages that show works, co-authors, and related topics.

Learn more

AI Enhancement of Linked Data

EBSCO believes that when used responsibly, AI can meaningfully enhance the research experience. By combining AI with advanced technologies and expert input, we aim to improve usability, ensure accuracy, and uphold the highest standards.

As part of this effort, we use AI to strengthen our linked data capabilities. For example, our existing Universal Subject Index (USI) has long used linked data to map synonyms across different subject authorities, helping users find relevant content even if they don't know the exact terminology. Building on this technology, we've implemented AI-assisted Natural Language Searching, which works in conjunction with the USI to better interpret user queries and return even more accurate results.

In addition, our Suggested Searches feature uses AI to analyze a user's query and recommend alternative or related searches, often guiding researchers toward deeper exploration of a specific subtopic.

**Responsible AI is central in our support
of research integrity and innovation.**

Explore our approach

EBSCO

Why is linked data important to libraries?

	Without linked data	With linked data
Discovery & Access	Researchers work harder to find content related to their research	Metadata is connected across datasets to reveal relationships between authors, works, and subjects, enhancing search and discovery
Metadata Quality	Data is disconnected and can require manual updates	Flexible, reusable data with dynamic authority control yields more consistent and up-to-date records
Interoperability	Library resources do not interconnect with other resources	Bridges between disparate vocabularies and ontologies unify access across libraries, archives, and museums
Open Knowledge	Closed systems are cut off from global ecosystem	Contributions to global open data ecosystems align with values of transparency and collaboration

How is EBSCO using Linked Data in our products and services?

Visit to find out