November 10–11, 2025

CONCERT Annual Conference, Taiwan

Bridging Intelligence: Practical AI applications in libraries

Axel Kaschte, PhD

Product Strategy Director, OCLC EMEA & APAC

We believe that AI is a game changer for libraries

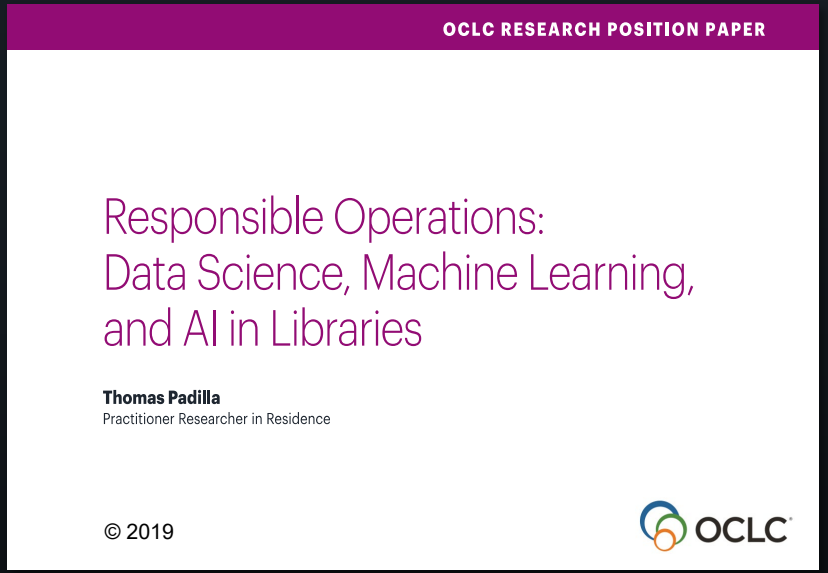
Internally - Making libraries more efficient
Externally - connect more deeply with their communities

OCLC

To realize this promise, OCLC has been on a six-year journey of responsible AI innovation

OCLCs 1st AI report in 2019

In Nov 2022, ChatGPT arrives…



OCLC Research Library Partnership:

AI for Metadata Workflows

- 2023-2024 OCLC / LIBER*
  "Building for the Future" series:
  Transformative technologies as an opportunity to elevate the role of the library

- 2024 OCLC RLP Metadata Managers Focus Group:
  "Getting ready for AI" discussions & report

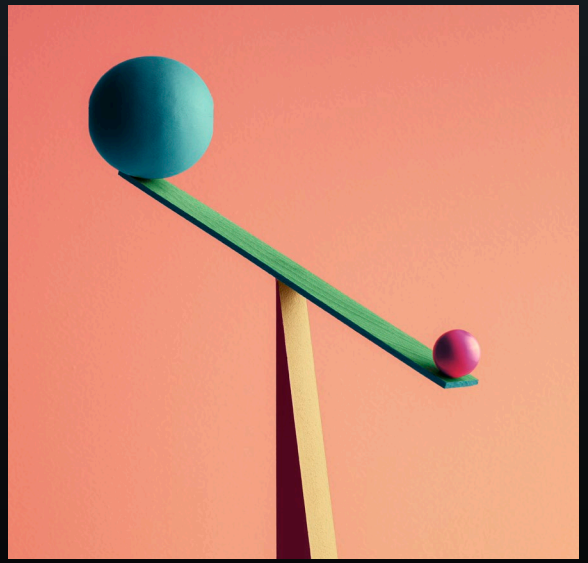
* LIBER is Europe's largest association of research libraries, representing 420+ university, national and research libraries.

OCLC

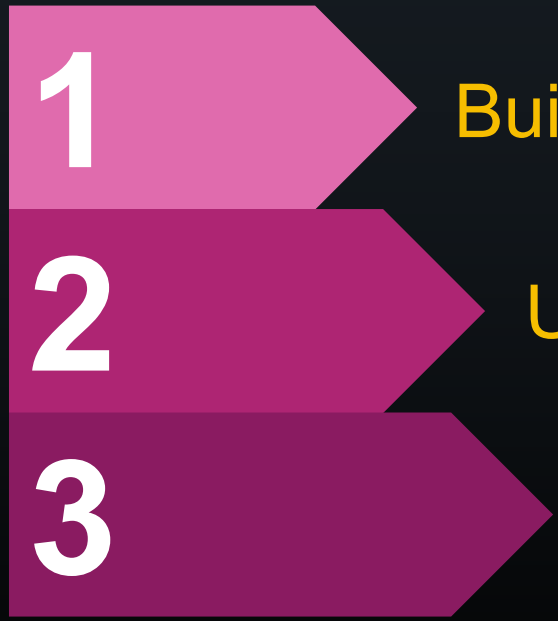
Building for the future:

Opportunities and Responsibilities

- Environmental impacts of significant energy usage
- Privacy of user data
- Use of copyrighted materials in LLMs
- Misinformation created by inaccuracies and hallucinations
- Risks of nefarious manipulations
- English language dominance in LLM models
- The ability to acquire relevant and usable information amidst intense information overload



Responsible AI Innovation



Build a value-based framework

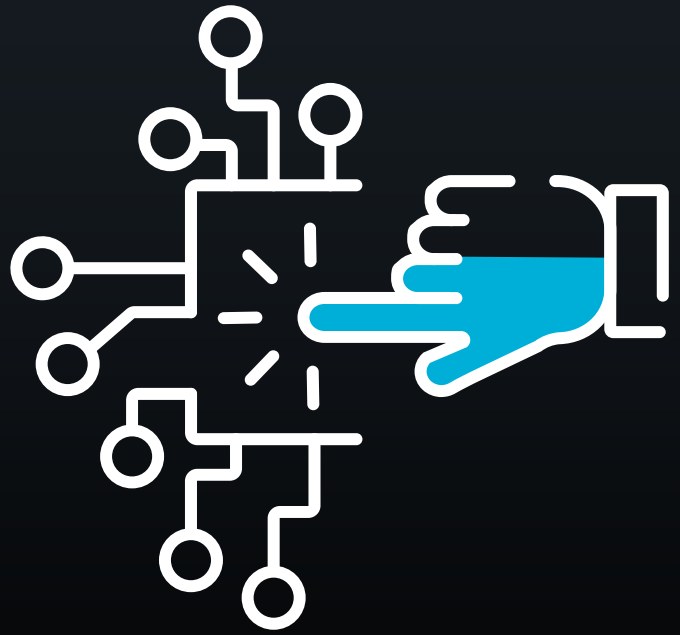
Understand libraries' uses for AI

Open exchange with the community

OCLC

AI and sustainability

We focus on sustainable AI by reducing energy use, optimizing resources, and prioritizing genuine impact.



What we're not doing with AI

OCLC does not hold, or process, the content of the publications it catalogs. None of the data in WorldCat has been used to train large language models like ChatGPT or Google Gemini.

OCLC's AI tools focus on services and data

Library services

Library data

Exploring AI Opportunities in Metadata Operations

OCLC

OCLC AI Initiatives (examples)

1. **Serendipitous discovery**: genAI-powered book recommender pilot in WorldCat.org and mobile app
2. **Scaling data quality operations**: Machine-learning deduplication pilots
3. **Supercharging cataloging workflows**: predictive classification pilot

Approach: focus on library values

Surface related titles using AI

User approach: Transparency

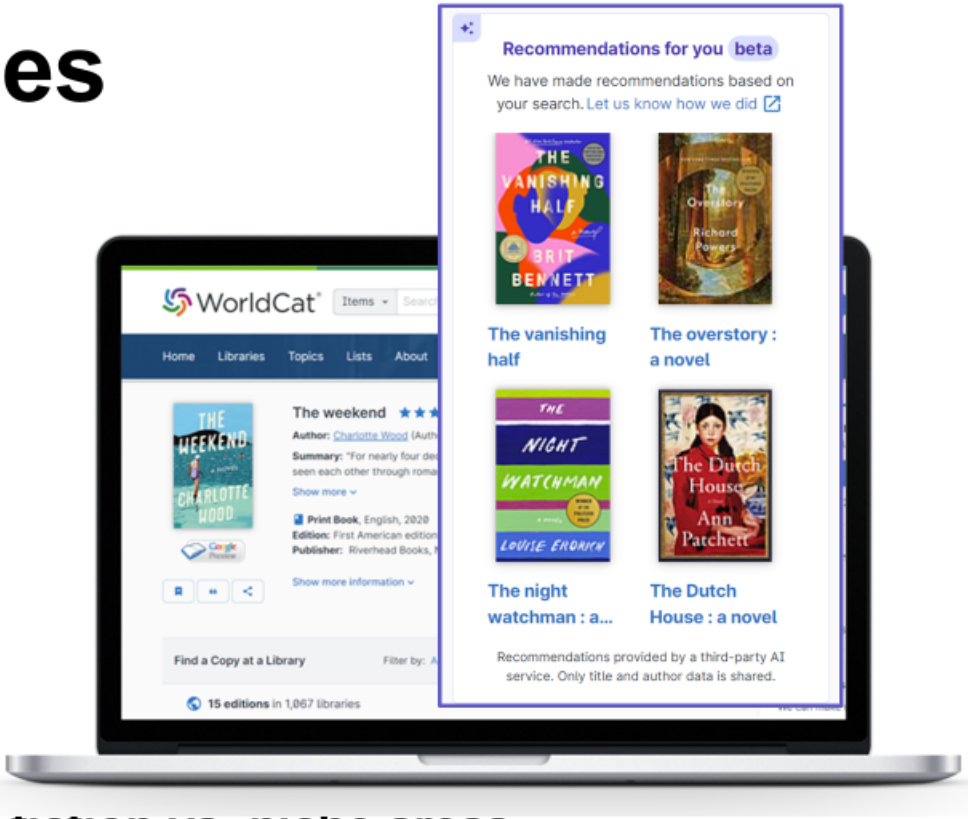
Technology approach: LLM

Hallucination check: WorldCat

Impact: 2.5 more titles viewed

Average pub'n date clicked: 2013

Works better for fiction, popular non-fiction vs. niche areas



Ex. 1: Serendipitous discovery

OCLC

High-quality metadata makes collections more discoverable

Quality improvement by merging duplicate records

32M+

new bibs created annually

20%

identified as duplicates using classic string matching

Very different records

describing the same instance

	Record 1	Fields	Record 2
☐	cam a 3	leader	cam a M
☐	504048055	c001	83566108
☐	900823s1900 ii 000 0 eng d	c008	931116s1900 ii 000 0 eng d
☐	$a 003545560 $2 Uk	d016	Missing Entry
☐	$a UKMGB $b eng $c UKMGB	d040	$a HLS $b eng $c HLS
☐	$a Suryanarain Row, Bangalore, $d 1856-	d100	$a Row, Banglore Suryanaraina.
☐	$a An Introduction to the Study of Astrology, in the light of physical sciences ... Third edition.	d245	$a An introduction to the study of astrology, $b in the light of physical sciences.
☐	Missing Entry	d250	$a 3d ed.
☐	$a Pp. ii. 57. Royal Victoria Press: $b Madras, $c 1900.	d260	$a Madras, $b K. Krishnaswami Ayar, $c 1900.
☐	$a 8°	d300	$a pages (4), ii, 57
☐	$a text $b txt $2 rdacontent	d336	$a text $b txt $2 rdacontent
☐	$a unmediated $b n $2 rdamedia	d337	$a unmediated $b n $2 rdamedia
☐	Missing Entry	d338	$a volume $b nc $2 rdacarrier
☐	Missing Entry	d776	$i Online version: $a Row, Banglore Suryanaraina. $t Introduction to the study of astrology. $b 3d ed. $d Madras, K. Krishnaswami Ayar, 1900 $w (OCoLC)902781618

Green box denotes identical strings in MARC record

Very similar records

describing different instances

Record 1	Fields	Record 2
cam a Ma	leader	cam a Ma
926921369	c001	926922235
960808s1958 gw 000 0 ger d	c008	960808s1958 gw 000 0 ger d
$a SHS $b eng $c SHS	d040	$a SHS $b eng $c SHS
$a Storm, Theodor, $d 1817-1888.	d100	$a Storm, Theodor, $d 1817-1888.
$a Works	d240	$a Works
$a Werke : $b Gesamtausgabe / $c hrsg. und eingeleitet von Hermann Engelhard. Bd. 1, Gedichte, Märchen, Novellen.	d245	$a Werke : $b Gesamtausgabe / $c hrsg. und eingeleitet von Hermann Engelhard. Bd. 2, Novellen.
$a Stuttgart : $b Cotta, $c 1958.	d260	$a Stuttgart : $b Cotta, $c 1958.
$a text $b txt $2 rdacontent	d336	$a text $b txt $2 rdacontent
$a unmediated $b n $2 rdamedia	d337	$a unmediated $b n $2 rdamedia
$a Engelhard, Hermann.	d700	$a Engelhard, Hermann.
$a Gedichte, Märchen, Novellen.	d740	$a Novellen.

Green box denotes identical strings in MARC records

Deduplication using machine learning

Combining different methods

Use **coded fields** to exclude records on general terms like language

Use **MinHash** to create a list of pairs of potential duplicates

Use **labelling tool** to manually label subset of pairs by experts

Use **Gradient Boosting model**

- Train model with labelled pairs
- Apply model to candidate pairs

New deduplication process—results

Conservative approach

Recognized 5.1M pairs of duplicates (in addition to existing process)

Confirmed 97% of GB model decisions with manual spot checks

Handled actual merging of duplicates with normal merge processes

Can unmerge, if false positive is found

Work in progress — more to come

Hanging Together

the OCLC Research blog

Scaling de-duplication in WorldCat

Balancing AI innovation with cataloging care

Photo by Bemal Rajapatirana on Unsplash

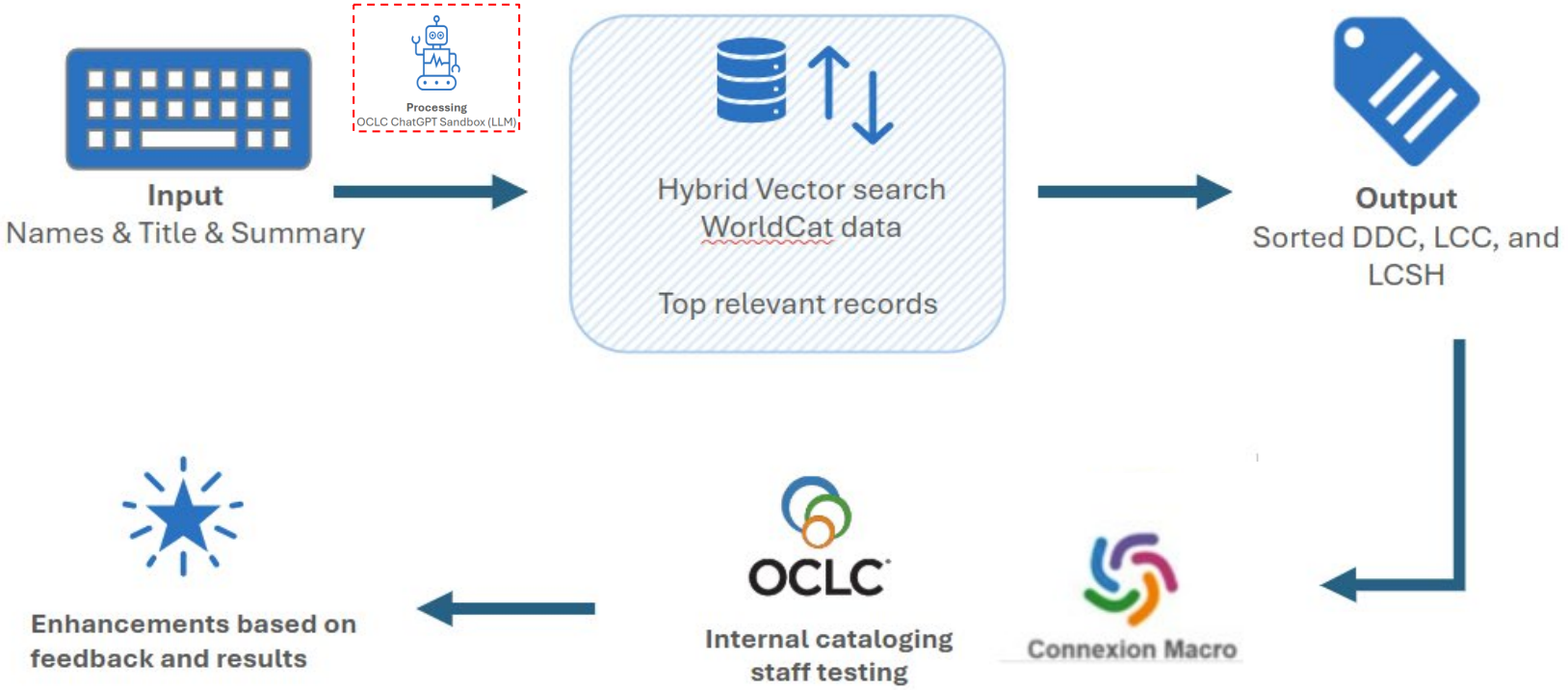
OCLC

OCLC

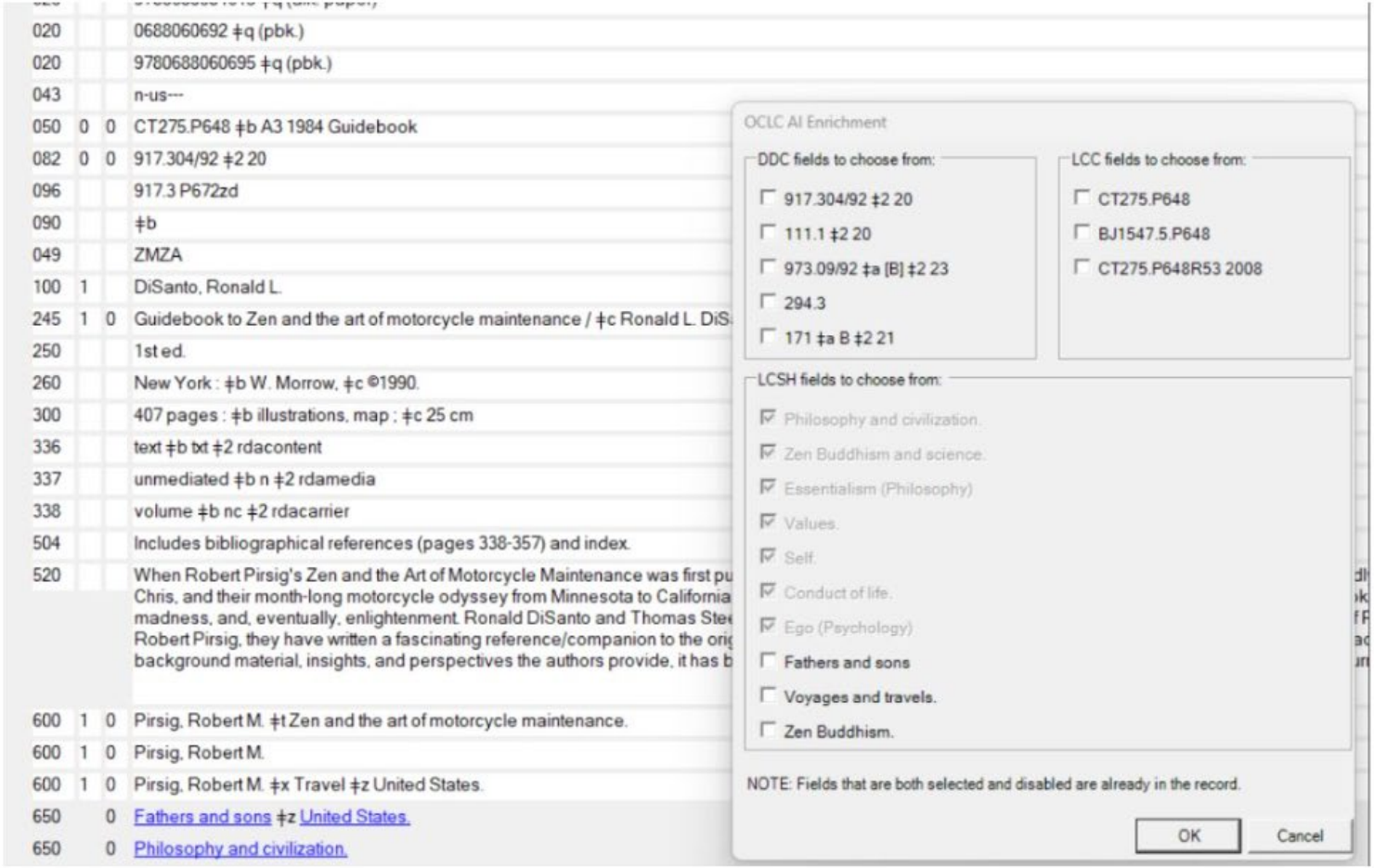
Enriching records with AI for better discovery

- Subject analysis and classification accounts for 70-80% of time for resource description
- Average time for record creation in MARC
  - Copy cataloging: avg. 20 minutes / record
  - Original cataloging: avg. 40 minutes / record
- Predicting subjects and classification can accelerate flow of materials to users

Iterative experimentation



Ex. 3: Supercharging cataloging workflows



Ex. 3: Supercharging cataloging workflows

Next steps

- Complete second round of testing with internal cataloging staff at OCLC
- Pilot with libraries using Connexion this quarter

Exploring AI Opportunities in Resource Sharing

OCLC

AI-driven global resource sharing for smarter, faster fulfillment

Automated request manager

Real-time availability

Smart lender strings

Express program

Intelligent timing

OCLC

AI-driven global resource sharing for smarter, faster fulfillment

AI significantly reduces average turn around times

Automated request manager

50% reduction

in wait times

OCLC

AI-driven global resource sharing for smarter, faster fulfillment

Items move faster through the network

Real-time availability

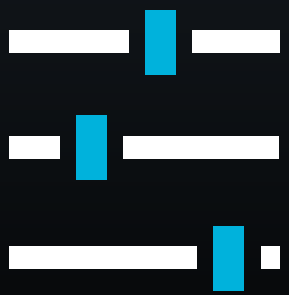
1,089,000+

Items were automatically moved to the next lender

OCLC

AI-driven global resource sharing for smarter, faster fulfillment

AI significantly reduces average turn around times



Smart lender strings

36% reduction

in wait times

OCLC

AI-driven global resource sharing for smarter, faster fulfillment

AI significantly reduces average turn around times

Express program

Fastest average

turnaround times in the industry–ten hours

OCLC

New! Intelligent timing

What is it?

Powered by AI and data—determines the actual times when digital requests are getting filled

---

Finds lenders across the globe that can fill a digital request the soonest

---

If you use automation to send copy requests to lenders, you are already using this

OCLC

Exploring AI Opportunities: Entity Metadata

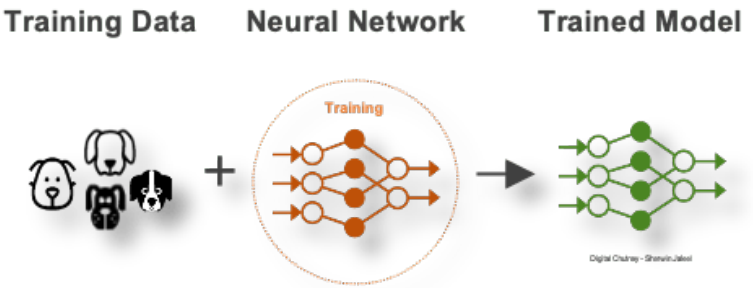
AI and Linked Data

Converging technologies

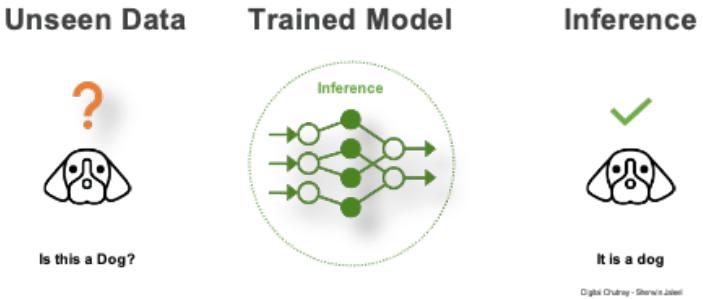
OCLC

Machine Learning – Two Steps

1) Training



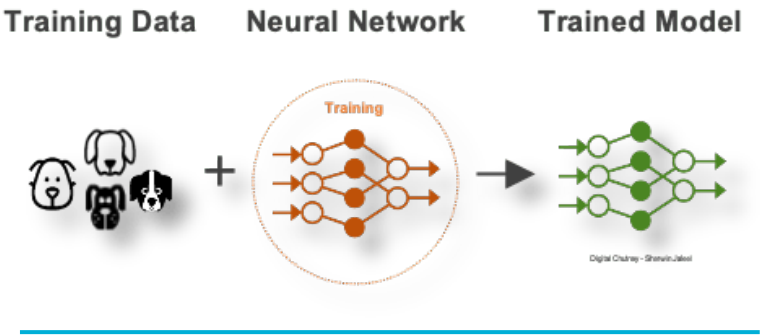
2) Inference



Pictures by Sherwin Jaleel (https://digitalchutney.blog/ml-training-inference)

Machine Learning - Challenges

- High data volume
- Data quality?
- Data consistency?



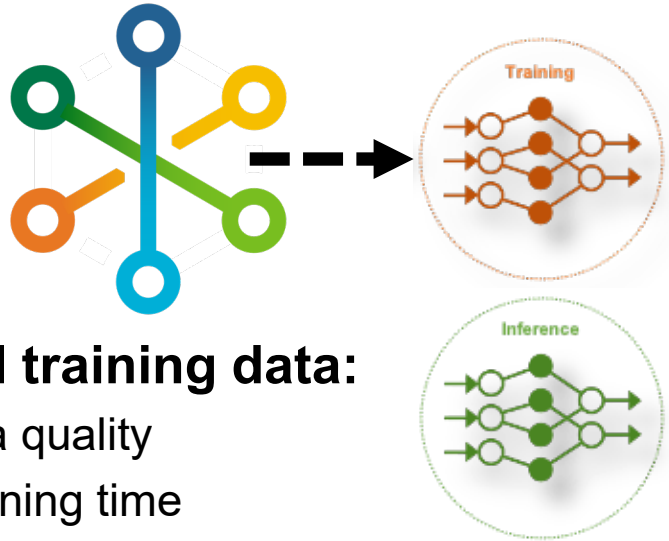
- Resource intensive
- Time consuming

- Resource intensive if many users
- Reasoning difficult
- Hallucinations

Pictures by Sherwin Jaleel (https://digitalchutney.blog/ml-training-inference)

OCLC

Linked Data improves AI



Structured training data:

- Higher data quality
- Shorter training time
- Smaller model size
- Much less resources

Structured inference context:

- Provides factual context, prevents AI hallucinations
- Traceable reasoning paths, explain AI conclusions
- Enables complex queries, beyond pattern recognition

GraphRAG

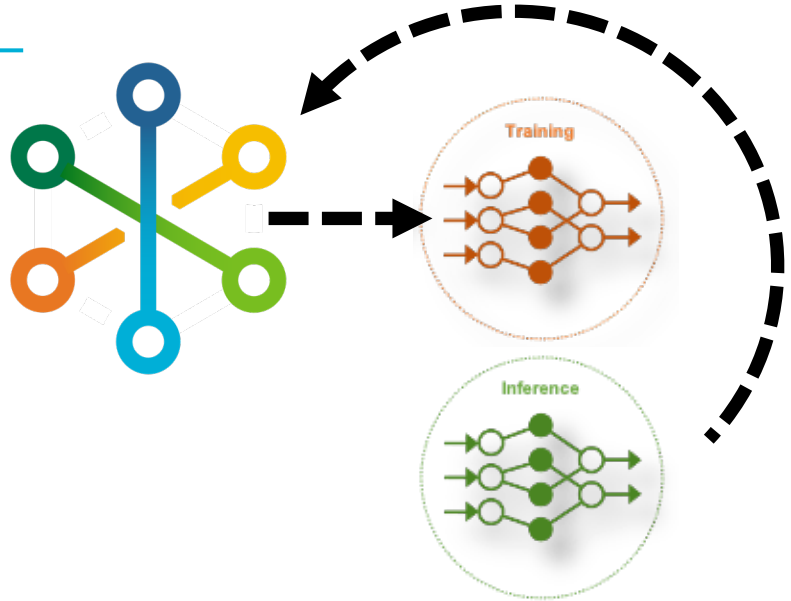
Graph Retrieval-Augmented Generation

OCLC

AI improves Linked Data

Continuous improvement cycle

- Update knowledge graphs based on new information
- Predict missing links in knowledge graphs
- Extracts entities and links from unstructured text
- Recognizes conceptual similarities across different ontologies



Linked data:
Expanding library reach

Connect library knowledge at scale

oc.lc/linkeddata

WorldCat® Entities

Things instead of Strings

Works

Persons

Places

Events

Organizations

entities.oclc.org

OCLC

OCLC Meridian

Create, curate, and connect linked data in WorldCat Entities

oc.lc/meridian

WorldCat® Entities

OCLC

Dewey Decimal Classification

Unlocks the power of Dewey numbers via the creation of URIs

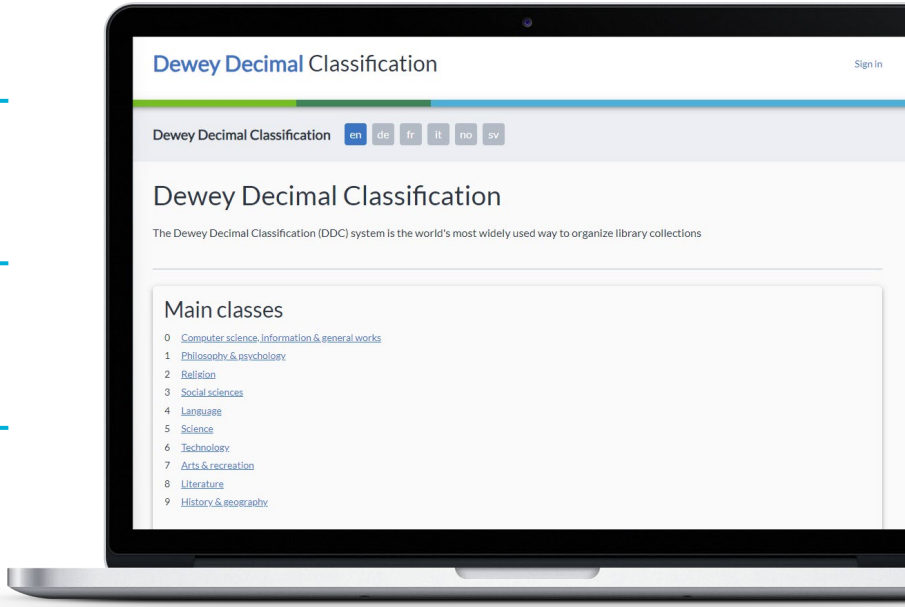
Forms the foundation for Concept entities

Provides subject classifications to Works entities in WorldCat Entities

Creates a layer of new relationships for knowledge graphs

Connects Dewey classifications to FAST topical headings

oc.lc/dewey-linked-data



OCLC

WorldCat® Entities in MARC records

Extend MARC cataloging towards knowledge work

Insert entity identifiers (URI) within existing workflows

650		0	Institutional repositories **$z** United States **$v** Case studies.
650		7	Institutional repositories. **$2** fast **$0** (OCoLC)fst01740486
700	1		Mixter, Jeffery, **$e** author. **$1** https://id.oclc.org/worldcat/entity/E39PBJtRJykHh8gwxXj3JQ4cyd
700	1		McMillian, Ciara.

OCLC

Strengthen and evolve the national communities
…AND at the same time
participate in global exchange
Syndeo™
Συνδέω (syndéo) = to connect
OCLC

National metadata management differs

National Metadata Management Practices

- National metadata management model
- National metadata format
- National metadata format extensions
- National collections and special sources
- National quality control

National Authority File Management

- Language specific authority files

National Metadata Hub

- Metadata aggregation
- Metadata mass changes on large scale

Syndeo™

National Metadata Management Practices

- National metadata management model
- National metadata format
- National metadata format extensions
- National collections and special sources
- National quality control

National Authority File Management

- Language specific authority files (e.g. Māori, French Canadian)

National Metadata Hub

- Metadata aggregation (e.g. NII Japan)
- Metadata mass changes on large scale

Tailor-made solution for each

OCLC

Syndeo for bespoke linked data

Create local, national, and regional entities

Build on local, national, and regional authority files

Allow language specific entities

Allow topically specific Eentities

A Syndeo library always retains full control of their linked data entities, while allowing all member libraries to benefit through WorldCat.

Connect to other entities → WorldCat® **Entities**

OCLC

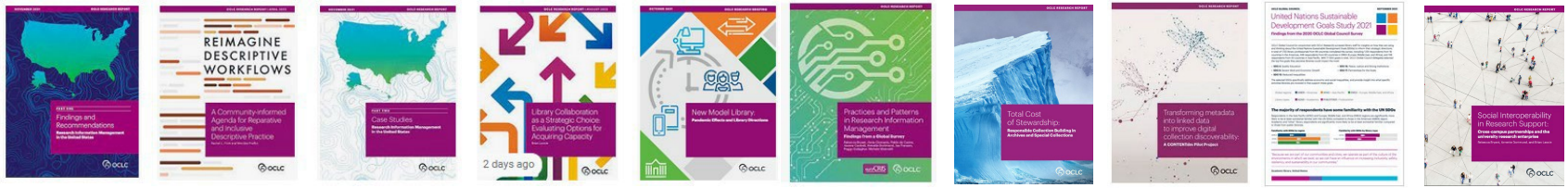
OCLC Research

OCLC Research

- 40 years: thought leadership on key issues for libraries
- 20 years: community programming to advance librarianship
  - WebJunction: professional development and skill-building for staff in public libraries
  - Research Library Partnership: peer-to-peer learning and leadership development for functional specialists in research libraries
- Combined capacity accelerates and scales library impact

Recent and Upcoming OCLC publications

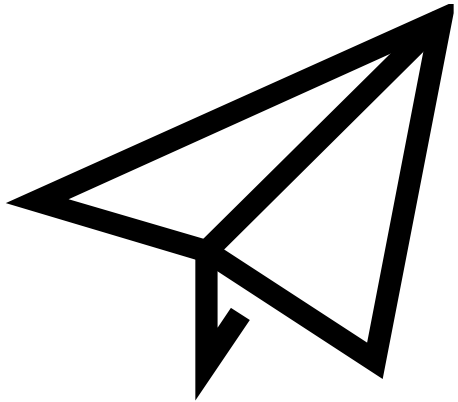
- **Migliorare la scoperta dell'open access per gli utenti delle biblioteche accademiche**
- Stewarding the Collective Collection: **An Analysis of Print Retention Data in the US and Canada**
- Stewarding the Collective Collection: **Data and Tools for Shared Print Workflows** – forthcoming
- **The Library Beyond the Library** - forthcoming



Do you want to receive the latest OCLC research reports, events, products and news?

Update your email preferences at:

oc.lc/mail



OCLC

Because what is known must be shared.®

Thank you

Axel Kaschte, PhD
Product Strategy Director,
EMEA & APAC
axel.kaschte@oclc.org

OCLC