

Leslie Harka

leslieharka@gmail.com | linkedin.com/in/lharka | Portfolio: lharka.com

PROFESSIONAL SUMMARY

MLIS graduate with 7+ years improving discovery, stewardship, and reuse of digital collections and research datasets. Experienced with repository platforms, metadata standards and linked open data, data curation aligned with FAIR and CARE, and computational methods for scaling metadata work. Collaborative teammate focused on staff mentoring, creating clear documentation, and implementing sustainable workflows.

EDUCATION

M.S. in Library and Information Science, University of Washington, 2025

Concentration: Metadata Design, Digital Archives, Records Management, and Knowledge Organization

M.S. in Community and Regional Planning, University of Oregon, 2021

Concentration: Community Archives and Digital Collections

B.S. in Psychology, University of Oregon, 2019

PROFESSIONAL EXPERIENCE

Research Collaborator, Web Archives Analysis Project

Mar 2025 – Present

Lab for Computing Cultural Heritage, University of Washington Information School

Seattle, WA

- Analyzing large-scale web archives to identify embedded data (e.g., spreadsheets embedded in PDFs)
- Applying AI/ML and multimodal approaches to improve discovery and metadata assessment
- Documenting workflows and contributing to publications on digital preservation and access

Graduate Staff Assistant

Sep 2022 – Jun 2025

iSchool GA Crew, University of Washington Information School

Seattle, WA

- Coordinated team workflows and project logistics for faculty research support, including tracking multiple concurrent requests, managing timelines and deliverables, and maintaining active communication with faculty and team members
- Trained and mentored new graduate assistants, creating documentation and training materials to ensure knowledge continuity and onboard team members efficiently
- Maintained faculty grant and publication records in Activity Insight database, ensuring accuracy and compliance
- Performed bibliographic and metadata analysis using scripting and query languages (Bash, Python, SQL, SPARQL); extracted document metadata from PDFs using Grobid and NLP techniques
- Managed collaborative writing projects for faculty publications with Git/GitHub and LaTeX (Overleaf, BibTeX)
- Supported research through literature reviews, citation management, data analysis, and publication/event graphics

Metadata Assistant

Nov 2017 – Aug 2022

Digital Scholarship Center, University of Oregon Libraries

Eugene, OR

- Assisted with large-scale metadata audits and remediation for thousands of digitized and born-digital materials during the Oregon Digital migration to Samvera/Hyrax repository to improve accuracy, interoperability, and discoverability
- Cataloged and indexed digital materials using linked data principles, controlled vocabularies, authority URIs, and spreadsheet-based workflows; managed controlled vocabularies in Oregon Digital's vocabulary manager
- Added new artist records to Getty ULAN, improving representation in UO's Oregon Percent for Art collection
- Collaborated with librarians and developers to improve repository workflows and metadata quality

Archival Processing Thomas Intern

Sep 2021 – Mar 2022

Special Collections and University Archives, University of Oregon Libraries

Eugene, OR

- Processed collections and created EAD-compliant finding aids in ArchivesSpace for physical and born-digital materials
- Cataloged records using Alma/Primo and OCLC Connexion; implemented access controls for restricted materials
- Applied preservation best practices and organized collections for long-term storage

Digital Collections Intern

Special Collections and University Archives, University of Oregon Libraries

Jun 2020 – Sep 2021

Eugene, OR

- Remediated metadata for 2,500 digitized photographs from the Doris Ulmann collection
- Designed a public-facing digital exhibit in Spotlight to showcase and contextualize the collection
- Collaborated with cross-departmental teams on metadata description, preservation, and access

Research Assistant

Social and Affective Neuroscience Lab, University of Oregon

Jun 2017 – Mar 2020

Eugene, OR

- Managed confidential health behavior research data for longitudinal study investigating behavior change outcomes, following IRB protocols and privacy standards similar to HIPAA
- Performed data quality control reviews to ensure consistency and accuracy in observational health data collection
- Coordinated research participant scheduling, documentation, and protocol compliance with attention to detail
- Provided participant support by explaining study procedures, obtaining informed consent, and addressing questions

PORTFOLIO & SELECTED WORK

Embedded Spreadsheets in Web Archives (2025-Present)

- Developing methods to detect spreadsheets in End of Term Web Archive PDFs using multimodal models and computer vision. Published tutorial on accessing parquet data on AWS S3. [\[Tutorial\]](#)
- *Tools: Python, SQL, AWS S3, Parquet, Web Archives, Data Extraction & Analysis*

ISMIR Metadata Analysis (2024-2025)

- Analyzed ISMIR conference metadata (2000-2024) for author affiliations and topics. Built Python pipeline using dblp Knowledge Graph and SPARQL endpoint to query and transform RDF data for bibliometric analysis. [\[Project Site\]](#)
- *Tools: Python, SPARQL, RDF, Knowledge Graphs, Data Extraction & Analysis*

Oregon Digital Migration (2021-2022)

- Metadata auditing and remediation for thousands of cultural heritage materials in the Oregon Digital platform relaunch (Samvera/Hyrax, Feb 2023). [\[Repository\]](#) [\[Live Site\]](#)
- *Tools: Samvera/Hyrax, Metadata Standards, Linked Data, Digital Repositories*

Archival Finding Aids (2021-2022)

- Processed collections and created EAD-compliant finding aids in ArchivesSpace for physical and born-digital materials. Selected examples: [\[1\]](#) [\[2\]](#) [\[3\]](#)
- *Tools: ArchivesSpace, EAD, Archival Description Standards*

Doris Ulmann Digital Exhibit (2020-2021)

- Remediated metadata for 2,500 Doris Ulmann photographs (early 20th-century American South) and created public Spotlight exhibit. [\[Collection\]](#) [\[Exhibit\]](#) [\[Blog Post\]](#)
- *Tools: Spotlight, Metadata Standards, Digital Archives, Exhibits*

SKILLS

Standards and Vocabularies: RDF, JSON-LD, Schema.org, DCAT, Dublin Core, MODS, MARC21, EAD, DACS, RDA, Wikidata, PIDs (DOI, ORCID), authorities and controlled vocabularies, FAIR and CARE

Repository and Library Systems: Samvera/Hyrax, ArchivesSpace, Alma/Primo, OCLC Connexion, Activity Insight

Programming and Query Languages: Python, SQL, SPARQL, Bash, XSLT, regular expressions

Data and Workflow Tools: Pandas, Jupyter Notebook, OpenRefine, Git/GitHub, REST APIs, AWS S3, Grobid (NLP)

Data Formats & Markup: JSON, XML, CSV, HTML, Markdown, Parquet

Digital Preservation: Records lifecycle, retention standards, preservation workflows, file format standards, quality control

Ethics and Privacy: IRB protocols, informed consent, sensitive data management (HIPAA-adjacent)

Project and Collaboration: Workflow management, process improvement, staff training and mentoring, documentation, cross-departmental collaboration, stakeholder engagement